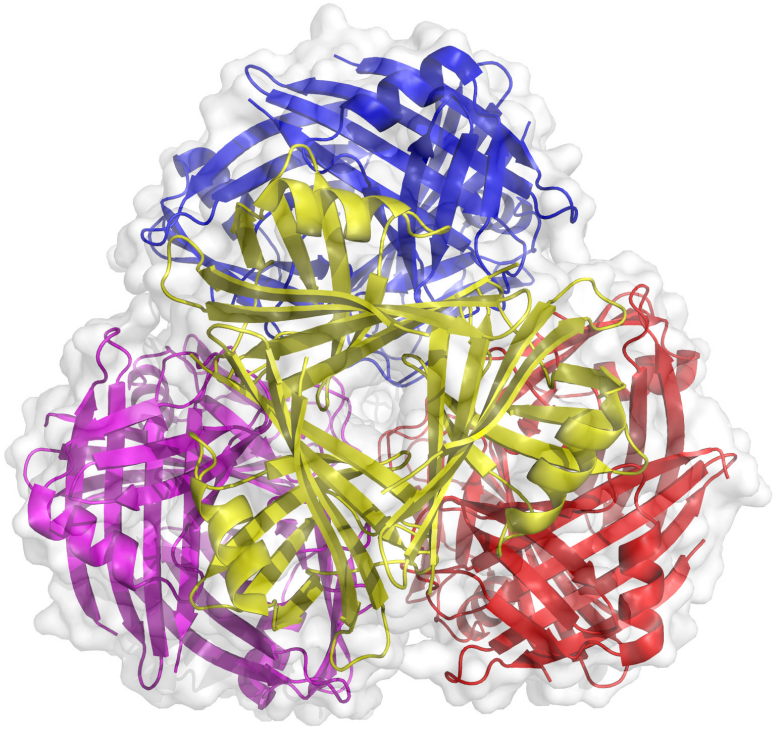
UNIVERSIDAD AUTÓNOMA DE MADRID

FACULTAD DE CIENCIAS

Departamento de Biología Molecular

Structural and functional characterisation of animal adenovirus capsid proteins



Abhimanyu Kumar Singh

Ph.D. Thesis

Madrid, 2014

UNIVERSIDAD AUTÓNOMA DE MADRID

FACULTAD DE CIENCIAS

Departamento de Biología Molecular

Structural and functional characterisation of animal adenovirus capsid proteins

Ph.D. Thesis

Abhimanyu Kumar Singh

Madrid, 2014

CENTRO NACIONAL DE BIOTECNOLOGIA

Obra Social Fundació "la Caixa"

CSIC Consejo Superior de Investigaciones Científicas

DECLARATION

I, Abhimanyu Kumar Singh, declare that the thesis entitled "Structural and functional characterisation of animal adenovirus capsid proteins" and the work presented in it, carried out at CNB-CSIC, Madrid, Spain under the guidance of Dr. Mark J. van Raaij, are my own. No part of this work has previously been submitted for a degree or any other qualification at this university or any other institution.

Abhimanyu Kumar Singh
Ph.D. candidate

Dr. Mark J. van Raaij
Thesis Director

Dr. Mauricio García Mateu
Thesis Tutor

Acknowledgements

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

First and foremost I offer my sincerest gratitude to my supervisor, Dr. Mark J. van Raaij, who has supported me throughout my thesis with his patience and knowledge whilst allowing me the room to work in my own way. He has been actively interested in my work and has always been available to advise me. His guidance helped me in all the time of research and writing of this thesis. One simply could not wish for a better or friendlier supervisor. Thank you Mark, for introducing me to the wonderful world of crystals and crystallography.

I would like to express my sincere thanks to La Caixa foundation for providing me the fellowship to carry out my doctoral research. I hope they will keep encouraging fellows with such initiatives in future as well for general betterment of science.

A good support system is important to surviving and staying sane during Ph.D. The members of the "structural biology of viral fibres" group have contributed immensely to my personal and professional time at CNB-CSIC. The group has been a source of friendships as well as good advice and collaboration. I would like to express my sincere thanks to Meritxell, Laura Cordoba and Carmela for their help and knowledge sharing during my initial days in the lab. I also wish to thanks Marta, Thanh and Laura Diaz for their friendly behaviour and making the lab atmosphere joyful. A special word of appreciation goes to Marta and Meritxell for helping me complete the Spanish summary part of the thesis and formatting assistance.

Completing this work would have been all the more difficult were it not for the support provided by the technical and non-technical staff of CNB-CSIC. I am indebted to them for their help.

I am deeply thankful to the people who have collaborated and supported me in my doctoral work: Mónika Z. Ballmann, Dr. Mária Benkő and Dr. Balázs Harrach (VMRI, Budapest, Hungary) for making available DNA samples and preparing some of the expression vectors, Dr. Carmen San Martín and Dr. Rosa Menéndez-Conejero (CNB-CSIC, Madrid, Spain) for providing genomic DNA of SnAdV-1, Dr. Michelle Kilcoyne and Prof. Lokesh Joshi (NUI, Galway, Ireland) for carrying out glycan array screenings, M. Álvaro Berbís and Dr. Jesús Barbero (CIB-CSIC, Madrid, Spain) for conducting STD-NMR experiments and Margarita Menéndez (IQFR-CSIC, Madrid, Spain) for performing ITC studies. Thanks to Mayte Martin and Leonor Kremer of the Protein Tools Unit at CNB-CSIC, Madrid for performing SPR

experiments and helping me interpret the results. I am grateful to Dr. Armando Albert for allowing me test crystals on his home X-ray source at IQFR-CSIC, Madrid. I also express my sincere thanks to the beamline scientists at ESRF (Grenoble, France) and ALBA (Barcelona, Spain) synchrotrons for their help in data collections and with whom I had the privilege to interact.

No work can be completed without a source of inspiration, and my family was that source of inspiration. I take this opportunity to express my profound gratitude to my parents and my siblings for their love, best wishes and continuous support.

Index

List of Figures

List of Tables

List of Abbreviations

BSA	Bovine Serum Albumin
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic Acid
GlcNAc	N-Acetylglucosamine
HAdV-2	Human Adenovirus 2
HAdV-19	Human Adenovirus 19
HAdV-37	Human Adenovirus 37
HAdV-5	Human Adenovirus 5
hCAR	Human coxsackievirus and adenovirus receptor
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	Isopropyl β-d-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
kb	kilobase
kDa	kiloDalton
LAdV-2	Lizard Adenovirus type 2
LB	Luria-Bertani culture media
MAD	Multi-wavelength Anomalous Dispersion
MAdV-2	Murine Adenovirus type 2
MES	2-(*N*-morpholino)ethanesulfonic acid
Ni-NTA	Nickel-Nitrilotriacetic Acid
O.D.	Optical Density
PBS	Phosphate Buffered Saline
PDB	Protein Data Bank
PEG	Poly Ethylene Glycol
r.m.s.d.	root mean square difference
SAD	Single-wavelength Anomalous Dispersion
SDS-PAGE	Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis
SL	Sialyl-lactose
SnAdV-1	Snake Adenovirus type 1
STD-NMR	Saturation Transfer Difference-Nuclear Magnetic Resonance
TAdV-3	Turkey Adenovirus type 3

TAE	Tris-acetate EDTA buffer
Tris	Tris(hydroxymethyl)aminomethane
v/v	volume/volume
w/v	weight/volume
w/w	weight/weight

Abstract

Adenoviruses are disease-causing agents, but are also explored as a model system to decipher mechanisms of molecular biology and used as vectors for gene and cancer therapy trials. Vectors wholly or partially based on animal adenoviruses may be interesting in this respect, both for avoiding pre-existing immunity in humans and the possibility of having novel cell-targeting properties. The capsid of the icosahedral, double-stranded DNA-containing adenovirus maintains the integrity of virus particle, but also has proteins involved in virus entry. The fibre is one such protein and is responsible for primary virus-host interaction. Hence, the homo-trimeric fibre protein is an important factor in adenoviral biology and to successfully exploit these viruses for medical applications.

In this thesis work, X-ray crystallographic structures of four animal adenovirus fibre heads have been determined, three of which belong to recently established genera of *Adenoviridae*, namely *Siadenovirus* and *Atadenovirus*. Despite being members of distinct genera and targeting various organisms, the structures revealed overall conserved topology of the central beta-sandwich in their head domains. However, other features of these fibre heads, such as their size, shape and the length of connecting loops show more diversity, which may, in turn, have an impact on their tropism.

Structures determined from two atadenoviruses have established their fibre heads as the smallest among all known adenovirus fibre head structures. Both the snake adenovirus 1 fibre head and the lizard adenovirus 2 short fibre heads are compact, with relatively short loops connecting their beta-strands. The presence of a prominent alpha-helix in their structures is also interesting.

Turkey adenovirus-3, which exists in avirulent and virulent forms, has a fibre head more similar to reovirus fibre heads than to those of other adenoviruses. The presence of a beta-hairpin insertion is unique to the structure. Moreover, the striking similarity of turkey adenovirus-3 fibre head with reovirus attachment proteins has strengthened the theory of an evolutionary link between adenoviruses and reoviruses. Sialic acid-containing carbohydrates were identified as *in vitro* ligands by NMR spectroscopy and calorimetry and co-crystal structures were successfully determined. The structure of ligand-bound turkey adenovirus 3 fibre head revealed the sialic acid binding site located on the side of the trimer.

Murine adenovirus 2, like most mammalian adenoviruses and all known human adenoviruses, belongs to the *Mastadenovirus* genus and could potentially be developed into a model system to study adenovirus pathogenicity. I have determined the high-resolution crystal

structure of the fibre head domain of murine adenovirus 2, along with the two directly preceding shaft domain repeats, which shows an overall resemblance to other adenovirus fibre heads of *Mastadenovirus* genus. Structural comparison with other fibre head structures, however, clearly shows variations that may have functional implications. N-acetylglucosamine, which is common component of glycosaminoglycans such as heparan sulfate and has a wide tissue distribution, has been identified as a ligand for murine adenovirus 2 fibre head in this thesis.

In summary, I present new structural and functional findings on distinct adenovirus fibre heads that will be useful in elucidating their tropism and biology. The new structures may also be useful in the design and development of animal adenovirus-based applications.

Introduction

1. Introduction

1.1 History and significance of adenovirus research

Adenoviruses were first isolated in 1953 from human pharyngeal tonsils (Rowe et al., 1953). The isolated agent was able to cause cytopathic changes and degradation of the adenoidal cells. Hilleman isolated a similar agent associated with acute respiratory illness from the tracheal phlegm and called it "adenoid degenerating agent" (Hilleman and Werner, 1954); later Huebner named them "adenoid-pharyngeal conjunctival and acute respiratory disease agents" (Huebner et al., 1954). Based on the observations of a common origin and close relatedness of the isolated agents, they were collectively named adenovirus (Enders, 1956). Further studies by different research groups firmly established the linkage of adenoviruses with acute respiratory illness and their frequent existence in a latent state in human tonsils and adenoids (Sohier et al., 1965). Furthermore, some adenovirus serotypes were found associated with infections of ocular tissues, such as adenovirus serotype 8 causing epidemic keratoconjunctivitis (Jawetz et al., 1957; Jawetz et al., 1955). Following the identification in humans, the first detection of adenovirus from non-human species was realized in cow (Klein, 1959) and in mouse (Hartley and Rowe, 1960). Later, they were isolated from and characterized in a number of further vertebrates species.

Due to their diverse occurrence and disease causing ability, several research groups focussed on adenovirus in the 1960s. During one such study, tissue culture fluid of human adenovirus serotype 12 (HAdV-12) was observed to successfully transform hamster cells (Trentin et al., 1962). Furthermore, it was shown that the HAdV-12 virus can also transform mice cells (Yabe et al., 1964). These two observations temporarily attributed adenovirus with a role in cancer progression. Subsequent research refuted this idea, but the intrinsic transformation ability of adenoviruses made them excellent tools to study the process of carcinogenesis. Splicing of introns was also observed for the first time in mRNAs of adenovirus serotype 2 (HAdV-2), during the late infection stage (Chow et al., 1977). This helped in establishing adenovirus as a model system to study gene expression in eukaryotes (Philipson, 1995).

During the past two decades, adenoviruses have been explored with two major objectives: to develop effective anti-virals against them and to adapt the virus into vector to be used in gene and cancer therapy. Additionally, some human adenovirus types have been used as oral vaccines against respiratory illnesses such as febrile acute respiratory disease. A vaccine

containing the avirulent form of turkey adenovirus 3 (TAdV-3) is widely used to prevent hemorrhagic enteritis in young turkeys.

Adenoviruses are associated with a number of pathogenic conditions such as epidemic keratoconjunctivitis, a severe ocular infection caused by some group D human adenoviruses (Martin et al., 2007). These viruses use cell surface sialic acid molecules as their attachment receptors and hence, inhibitors based on sialic acid are currently being developed against these viruses (Spjut et al., 2011). Although these conditions are usually self-limiting and infection generally induces sufficiently strong adaptive and protective immunity, patients have to suffer due to associated clinical manifestations. Moreover, in the case of immunosuppressed patients, infection can lead to severe clinical complications such as organ failure (Kinchington et al., 2005). Despite all efforts, there are currently no anti-viral agents against adenovirus for systemic or topical administration (Kinchington et al., 2005).

As far as vector development is concerned, the decade of the 90s has seen significant efforts in designing vector systems based on adenoviruses, thanks to their distinctive features such as a high potential trans-gene capacity, possibility of recombinant growth up to high titer and efficient transduction (McConnell and Imperiale, 2004; Bachtarzi et al., 2008). Patients may have pre-established immunity against human adenoviruses, so use of non-human adenoviruses might be advantageous, as they may be less immunogenic and may have novel receptor-binding properties (Loser et al., 2002).

1.2 Genome organisation

The adenovirus genome is a linear, non-segmented, double stranded DNA with characteristic inverted terminal repeats (ITR) and a covalently attached terminal protein at the 5'-end of each strand. Depending on genus and adenovirus type, the genome is between 26 and 48 kb in length, along with variations in features such as length of the ITR and degree of sequence conservation, however, these variations are mainly observed in the terminal regions and the central part of the genome is generally conserved (Figure 1; (Mautner et al., 1995; Harrach, 2011)). Variations in total G+C content of genomes have also been observed among different adenovirus types.

Many of the adenoviral RNA polymerase II transcription units are poly-cistronic. The products of these units undergo alternative splicing to yield different products (Berget et al., 1977). The transcriptional units were named according to whether their mRNA products accumulate in the early (E1A, E1B, E2A, E2B, E3, and E4) or late phase (L1 to L5) of infection (Flint et al., 1975; Russell, 2000). Furthermore, the gene products expressed in the

early phase are associated with viral DNA replication, while the production of virus structural components and the encapsulation and maturation of virus particles takes place in the late phase (Russell, 2000).

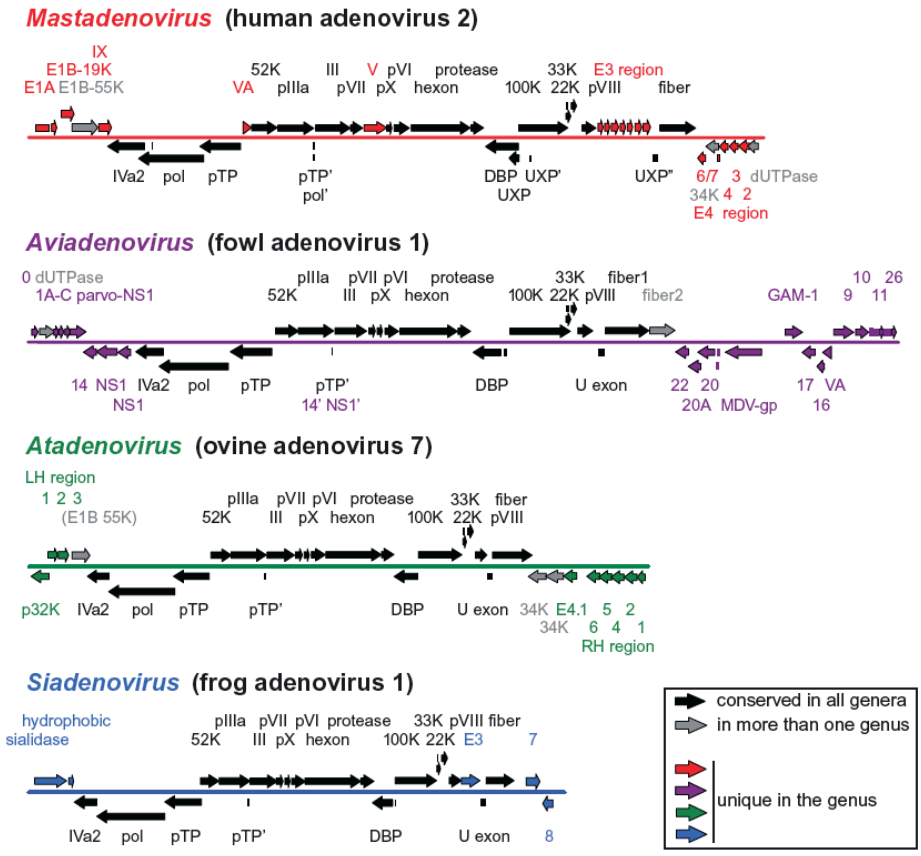


Figure 1: Genome organization of representative species of four adenovirus genera. Black arrows show conserved genes; grey arrow indicates genes that are present in more than one genus; colored arrows show genus-specific genes. Genome map for the fifth genus *Ichtadenovirus* is not yet available. Modified from (Harrach, 2011).

1.3 Adenovirus virion architecture

The adenovirus virion is around 90 nm in diameter and displays icosahedral symmetry with a triangulation number of 25 (T=25). Each non-enveloped, mature particle has an outer capsid and an inner DNA containing central core that accounts for a combined mass of approximately 150 x 10^6 Da (Vellinga et al., 2005; Shenk, 1996). Proteins account for 87% of the total mass whereas a single DNA molecule contributes the remaining 13%. Trace amount of N-acetyl-glucosamine is also present as an added moiety to the fibre protein, which is the only adenoviral protein undergoing glycosylation (Beach et al., 2009a).

The faces of adenovirus capsid are made up of the hexon protein, while the penton base forms the vertices. Into each vertex, a trimeric fibre is inserted. Besides these, there are some minor structural proteins, which are necessary for assembly and function as capsid stabilizers

(San Martin, 2012). In atadenoviruses, the capsid additionally contains two genus specific proteins, p32 and LH3 (Löser, 1999). Figure 2 shows a schematic representation of different polypeptides and their location in the viral capsid. Capsid associated proteins of adenoviruses can be divided in two categories:

1 Major capsid proteins

2 Minor capsid proteins

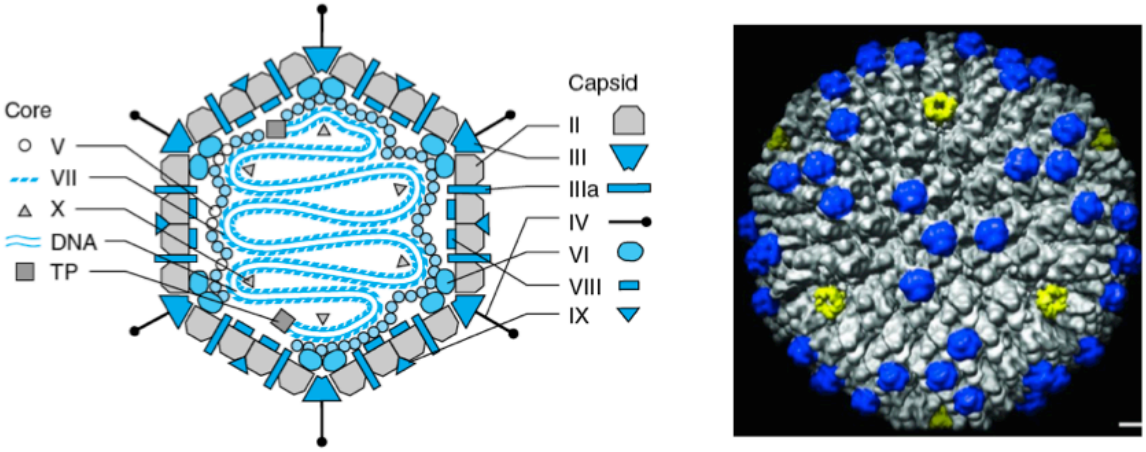


Figure 2: Location of adenovirus proteins (A) Schematic diagram of human adenovirus (HAdV) showing the eleven capsid and core proteins alongwith the viral DNA. Modified from Stewart and Burnett (1993). (B) Cryo-EM reconstruction of ovine adenovirus, a member of *Atadenovirus* genus. The genus specific LH3 protein is shown in blue and pentons are highlighted in yellow (Pantelic et al., 2008).

1.3.1 Major capsid proteins

Hexon: The hexon or polypeptide II is the most abundant and primary constituting unit of the adenovirus capsid. Twelve copies of the hexon trimer form each of the 20 icosahedral facets of the virion; the 720 total copies of the hexon monomer account for 63% of the virion protein mass (van Oostrum and Burnett, 1985; Rux and Burnett, 2004). It is worth mentioning that the hexon was the first animal virus protein to be crystallized (Pereira et al., 1968).

A single hexon subunit is composed of two anti-parallel, eight stranded beta-barrels, which are localized towards the bottom of the structure. In the trimeric assembly, the 3-fold repetition of these beta-barrels gives the pseudo-hexagonal appearance to isolated hexons (Rux and Burnett, 2004). Some of the loops connecting the strands reach out to the top of the trimer and interact with loops of adjoining subunits. These inter-monomeric interactions confer remarkable stability to the trimer.

Apart from inter-monomeric interactions within a hexon, the terminal regions of the hexon monomer have also been implicated in facilitating interactions between two hexons and between a hexon and minor coat proteins IIIa and VIII (San Martin, 2012). These interactions further contribute towards capsid integrity. Some studies have also shown that the outer regions of hexon are involved in receptor binding and virus internalization (Waddington et al., 2008; Kalyuzhniy et al., 2008).

Penton base: The penton base is a pentameric assembly of polypeptide III. Together with the trimeric fibre protein it constitutes the vertex capsomers or pentons of adenoviruses. The penton base exhibits a high level of amino acid sequence conservation even among human and non-human adenoviruses, with only the exposed hypervariable regions showing high sequence divergence (Zubieta et al., 2005). The atomic structure of the penton base revealed a 30 Å hollow central core. The N-terminal region of each monomer is directed towards the viral core, while interacting with the minor capsid stabilizing protein IIIa (Liu et al., 2010). An alteration in the interacting region of IIIa results in premature loss of pentons from the capsid (San Martin et al., 2008). At the capsid surface, the jellyroll arrangement of each monomer provides five equivalent interfaces that can accommodate incoming fibre tail monomers (Zubieta et al., 2005).

The vertex capsomers are responsible for the crucial cell entry process. The distal C-terminal part of the fibre recognizes specific cell surface receptors and attaches the virus to host cell in a convenient orientation (Bergelson et al., 1997; Arnberg et al., 2000). This is then followed by a secondary interaction between the penton base and cell surface integrins, promoting viral internalization via endocytosis (Wickham et al., 1993; Li et al., 2001; Soudais et al., 2000; Bangari and Mittal, 2005). Although the secondary interaction of integrins with the penton RGD motifs greatly enhances the efficiency of cellular entry, absence of RGD motifs doesn't prevent viral infection (Albinsson and Kidd, 1999; Hidaka et al., 1999). In the absence of other virion components, pentons can assemble as dodecamers and enter into the cell by endocytosis (Fender et al., 1997; Zubieta et al., 2005).

Fibre: The adenovirus fibre is a homo-trimeric, long, surface exposed protein that is attached to the penton base at each of the capsid vertices and protrudes away from the viral surface (Figure 3A). Its primary function is to mediate virus-host attachment through interactions with cell surface receptors such as CAR, sialic acid, desmoglein 2 or CD46/80/86 (Bergelson et al., 1997; Arnberg et al., 2000; Wang et al., 2011; Segerman et al., 2003; Short et al., 2006). The fibre is one of the major contributors to viral antigenecity, along with the hexon and the penton base. It has also been linked to viral maturation and intracellular

trafficking (Legrand et al., 1999; Harrach, 2011). Generally, the number of fibres per vertex is one, however, there are exceptions (Hess et al., 1995). Historically, the fibre was the second adenovirus structural protein to be crystallized (Mautner and Pereira, 1971), however, successful structure determination took another 23 years (Xia et al., 1994).

Structurally, the fibre can be divided into three domains; an N-terminal virus attachment or tail domain, a central shaft domain and a distal C-terminal globular head or knob domain (Figure 3A). The tail domain contains well-conserved motifs (Chroboczek J, 1995), of which one (ENPVYPY) anchors the tail to the penton base (Zubieta et al., 2005). The central shaft domain contains triple beta-spiral repeats with the consensus sequence XXφXφXφX-T-X#φXφXX-L (X is any amino acid, # is typically a proline or glycine, φ is a hydrophobic residue, T is a turn region and L a loop region). Amino acid insertions are tolerated in the loop region and, to a lesser extent, in the turn (Figure 3B; van Raaij et al. (1999b)). The length of the shaft domain is quite variable between adenovirus types, depending upon the number of triple beta-spiral repeats present (Nicklin et al., 2005). For example, the human adenovirus serotype 3 (HAdV-3) fibre shaft has only six triple beta-spiral repeats whereas serotypes 2 and 5 contain 23 (Signas et al., 1985; Green et al., 1983). This variation may affect viral infectivity and receptor specificity (Shayakhmetov and Lieber, 2000). Among different adenoviruses, the fibre head domain varies significantly in sequence. In contrast, their structures show conservation of topology, while major variations are restricted to connecting loops (Figure 3B; (Xia et al., 1994; van Raaij et al., 1999b; Guardado-Calvo et al., 2007; El Bakkouri et al., 2008; Guardado-Calvo et al., 2010)). This conservation extends to reovirus fibres and certain bacteriophage receptor-binding proteins as well (Chappell et al., 2002; Spinelli et al., 2006). Each monomer of the adenovirus fibre head trimer has a central eight-stranded beta-sandwich, the strands of which are connected through loops of varying lengths (Xia et al., 1994).

The globular nature of the fibre head allows its sides and top surface for engaging host receptors and, at the same time, the shaft domain provides reach and flexibility (Nicklin et al., 2005). Given its crucial role in receptor recognition process, the fibre is regarded as an important determinant of adenoviral tropism, which makes it an object of study in adenovirus-based gene vector systems (Nicklin et al., 2005; Noureddini and Curiel, 2005).

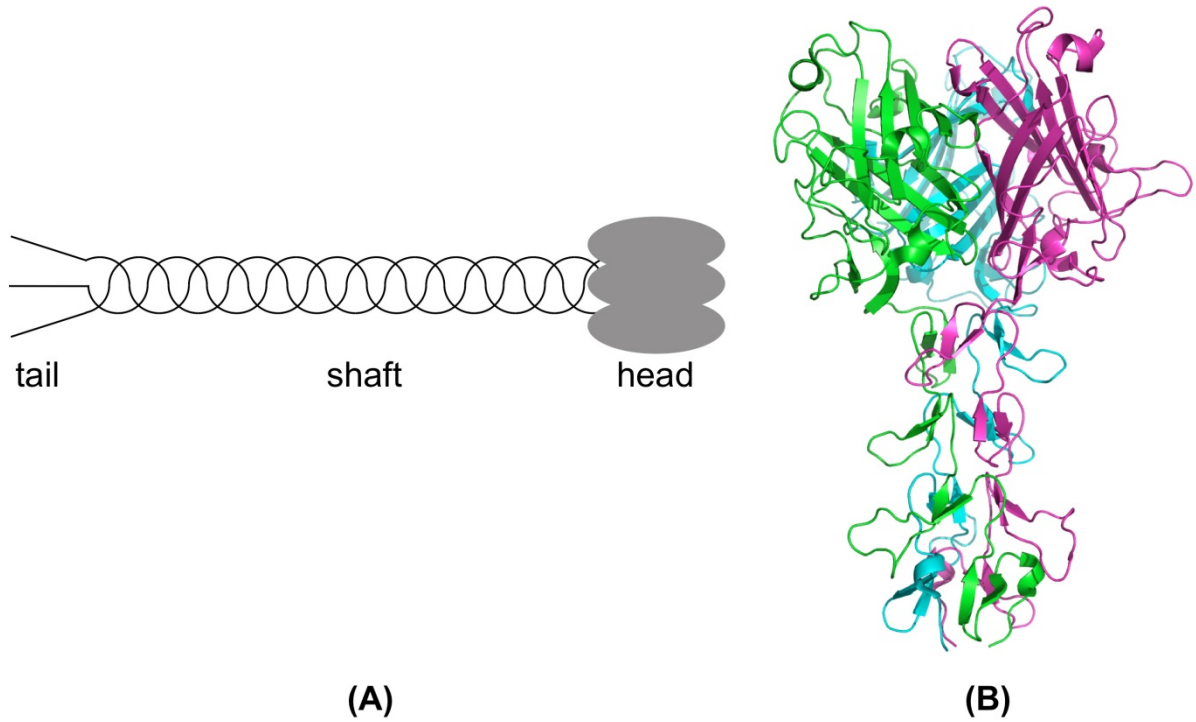


Figure 3: Adenovirus fibre (A) Schematic representation highlighting the N-terminal tail, the central beta-spiral repeat-containing shaft and the C-terminal receptor-binding head domain. (B) Crystal structure of the HAdV-2 fibre head domain with a portion of the shaft region. Monomers are shown in different colours (van Raaij et al., 1999b).

LH3: LH3 is a trimeric protein, of which each monomer weighs 42 kDa, and is present in members of *Atadenovirus* genus only. It is one of the two genus-specific proteins of atadenoviruses, the other being p32. Initially it was thought to be a homologue of the E1B55 kDa protein, which is a non-structural regulatory in protein human adenoviruses (Vrati et al., 1996). LH3 is a capsid component in ovine adenovirus, with approximately 1500 copies (i.e. 500 trimers) per virion (Gorman et al., 2005). Later, a cryo-EM reconstruction at 10 Å revealed the location of LH3 on the outside of the capsid, forming prominent surface protrusions (Figure 2B). It occupies a position equivalent to that of human adenovirus capsid protein IX (Ghosh-Choudhury et al., 1987; Pantelic et al., 2008) and may be involved in capsid stabilization and/or receptor interaction (Löser, 1999; Pantelic et al., 2008).

1.3.2 Minor capsid proteins

Proteins IIIa, VI, VIII and IX are four minor capsid components in human adenoviruses. Their contributions towards virus structure and function is not completely understood (Vellinga et al., 2005). *Protein IIIa* is a 63.5 kDa protein and is generated by the N-terminal cleavage of a 67 kDa precursor molecule (Vellinga et al., 2005). Protein IIIa acts like a cementing material below the vertex to hold adjoining hexons together and thus maintains the stability of capsid facets (Liu et al., 2010). This crucial aspect is reflected in assembly

malfunctions of virons in case of a defective or non-functional protein IIIa (Chroboczek et al., 1986).

Protein VI is a 22 kDa protein. It holds peripentonal hexons together and connects individual hexons trimers to the double stranded DNA (Figure 2A; (Wodrich et al., 2003; Russell and Precious, 1982; Stewart et al., 1993)). Free protein VI, released during pH-dependent viral disassembly, destabilizes the endosomal membrane, facilitating viral escape (Wiethoff et al., 2005; Maier et al., 2010). Newly synthesized hexons, which are bound to protein VI, are transported to the nucleus through the interaction of a C-terminal nuclear localization signal sequence of protein VI with the importin protein (Wodrich et al., 2003).

Protein VIII is the least characterised of all minor capsid proteins and exist in a dimeric form inside the viral capsid (Figure 2A; Vellinga et al. (2005)). Mutational analysis has shown its contribution in maintaining virion stability and integrity (Liu et al., 1985). In case of porcine adenovirus 3, interaction of protein VIII with the packaging motor protein IVa2 has been observed, which points to its probable role in genome packaging (Singh et al., 2005).

Protein IX is a genus-specific component of mastadenoviruses and is not present in any other genera of the *Adenoviridae* family (Figure 2A). It is the smallest of all minor coat proteins with a size of 14.3 kDa and cements hexon-hexon interactions (Everitt and Philipson, 1974; Vellinga et al., 2005; San Martin, 2012). Protein IX is not essential for virus assembly but IX-deletion mutants exhibit poor thermo-stability (Colby and Shenk, 1981).

The central core region of the adenovirus virion contains polypeptides V, VII, μ, IVa2, p23 and the terminal protein (TP) (Berk, 2007; Echavarría, 2004). The first three of these proteins are basic in nature and interact with the DNA (Anderson et al., 1989; Hosokawa and Sung, 1976; Russell et al., 1968). Protein V acts as the linker between the core and the capsid proteins VI and the penton base (Everitt et al., 1975; Matthews and Russell, 1998), whereas pVII and μ facilitate viral DNA condensation (Chatterjee et al., 1986; Rux and Burnett, 2004). TP covalently attaches to the viral DNA termini *via* a phosphodiester linkage, circularizes it and acts as a primer for DNA replication (Rekosh et al., 1977). Polypeptide IVa2 is a multifaceted protein, which plays roles in transcription regulation and DNA packaging (Gustin et al., 1996; Zhang and Arcos, 2005). Protein 23 is the principal viral protease, responsible for the cleavage of several capsid and core proteins (Anderson, 1990).

1.4 Taxonomic classification

Classification of early human adenoviruses was done based on their ability to agglutinate human and animal red blood cells (Rosen, 1958; Hierholzer, 1973). With the advent of more

advanced methodologies this was replaced by other, more precise criteria such as virion polypeptide profiles, DNA restriction endonuclease patterns and DNA hybridization analyses (Wadell, 1979; Wadell et al., 1980; Wigand et al., 1980; Garon et al., 1973; Green et al., 1979). Presently, there are 54 known serotypes of adenoviruses infecting humans, which are grouped under seven species (Martin et al., 2007), A-G, and have been associated with different tissue tropisms (Table 1).

Phylogenetic distance, genomic distance, nucleotide composition and cross-neutralization has also been used for classification (Harrach, 2011) and now, the family *Adenoviridae* consists of five recognized genera; *Mastadenovirus* (infecting mammals), *Aviadenovirus* (infecting birds), *Atadenovirus* (infecting ruminants as well as reptiles and birds, *Siadenovirus* (infecting birds and frog) and *Ichtadenovirus* (infecting fish) (Figure 4; Harrach (2011)).

Table 1: Human adenovirus serotypes and associated infections

Species	Serotype	Infection
A	12, 18, 31	Intestine
B	3, 7, 16, 21, 50 11, 14, 34, 35	Respiratory tract, eye Respiratory and/or urinary tract, eye
C	1, 2, 5, 6	Respiratory tract
D	8-10, 13, 15, 17, 19, 20, 22-30, 32, 33, 36, 37, 38, 39, 42-49, 51, 53, 54	Eye, intestine
E	4	Respiratory tract, eye
F	40, 41	Intestine
G	52	Intestine

Modified from Cupelli and Stehle (2011)

The name Atadenovirus found its origin in the fact that the first characterised atadenoviruses had a strong bias towards A+T in their genomes (Benko and Harrach, 1998). Adenoviruses from the *Atadenovirus* genus have been isolated from squamate reptile hosts, ruminants and birds and have a characteristic gene organization and capsid morphology.

The origin of Siadenoviruses is not clear and they have been isolated from one amphibian host, from Sulawesi turtles and from several species of birds (Davison et al., 2000; Harrach, 2011; Kovacs and Benko, 2011; Park et al., 2012). The genus takes its name from a specific open reading frame with high sequence similarity to bacterial sialidase proteins (Davison et al., 2000), although the functional significance of this gene is not known.

Only one Ichtadenovirus has so far been discovered, white sturgeon adenovirus 1, with the longest known adenovirus genome of around 48 Kb (Harrach, 2011).

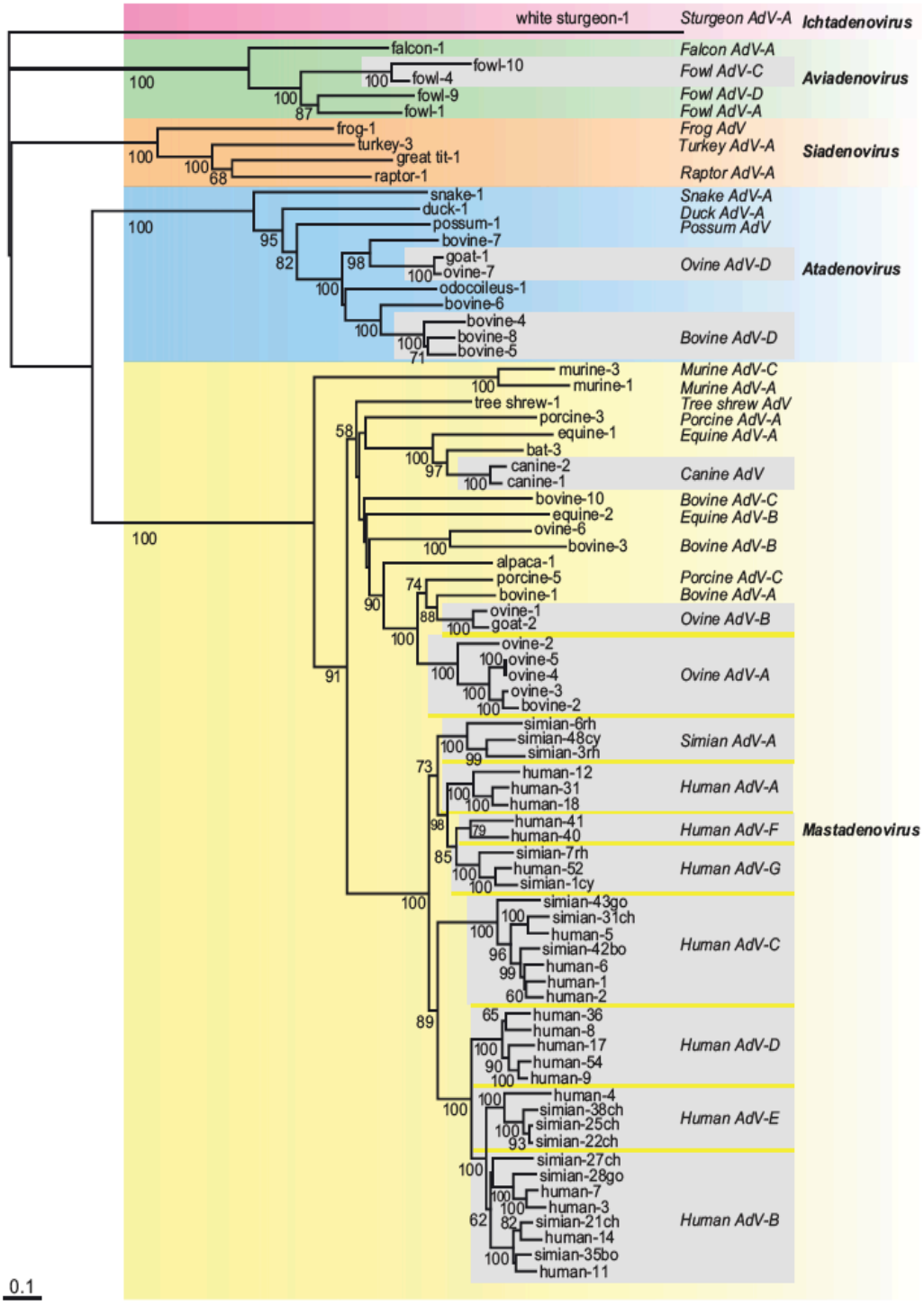


Figure 4: Phylogenetic tree of adenoviruses based on a distance matrix analysis of hexon amino acid (Harrach, 2011).

1.5 Overview of the adenovirus life cycle

Most of the studies on the adenovirus life cycle have been carried out on human adenovirus serotypes 2 (HAdV-2) and 5 (HAdV-5), due to the ease of their *in vitro* production (Berk, 2007). Figure 5 shows a schematic representation of adenovirus entry into the cell and its subsequent processing. Most of these steps are likely to be general to adenovirus infection, although some differences with the animal adenoviruses studied in this thesis are also likely to exist.

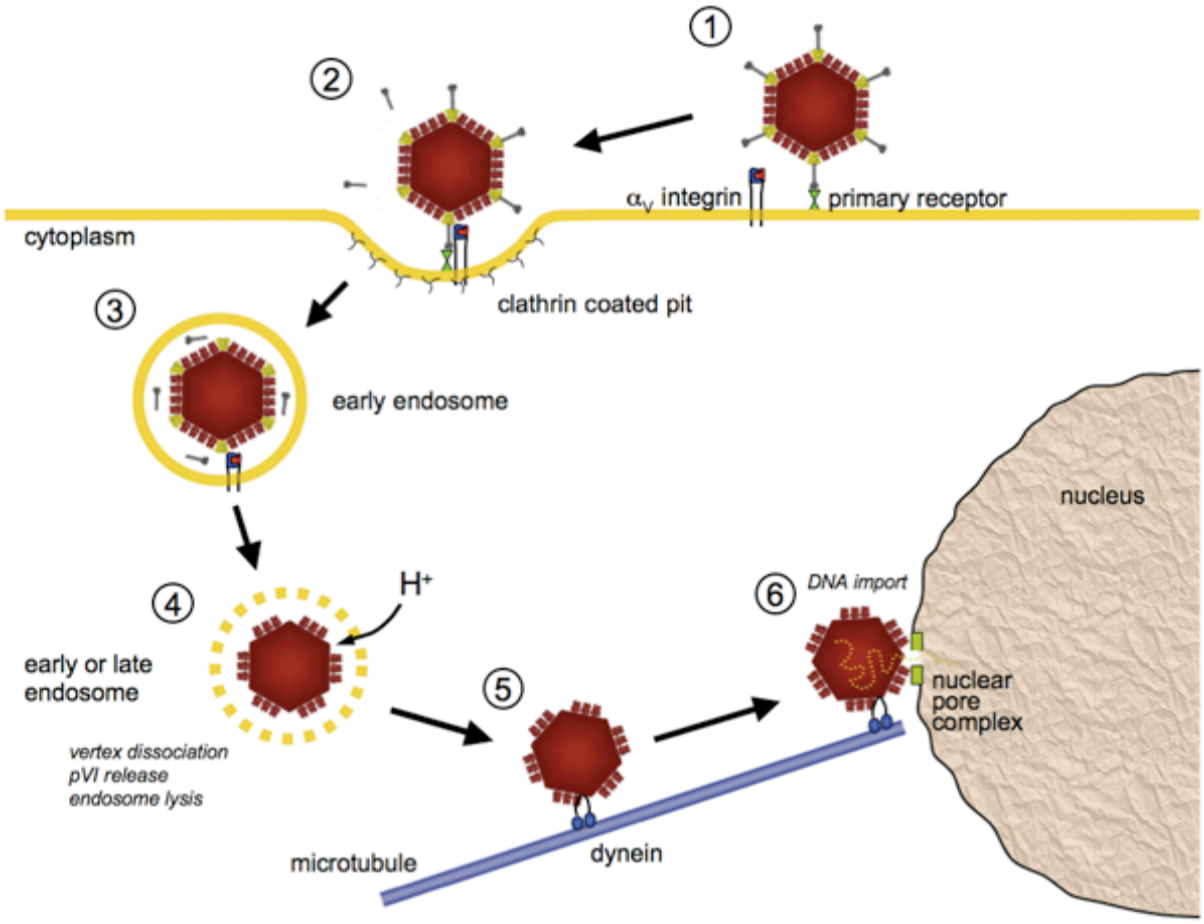


Figure 5: Adenovirus cellular entry (1) Virus attachment to cellular receptor (2) Internalization via endocytosis. (3) Partial disassembly of the virion inside early endosome (4) Interaction of pVI with endosomal membrane leading to membrane disruption and capsid release. (5) Transport of capsid to nuclear periphery where it docks on nuclear pore complex. (6) Complete capsid disassembly and DNA release into the nuclear pore complex. Taken from Nemerow et al. (2009).

The first step of adenovirus infection cycle begins with the attachment of virus particles to specific host cell receptors. The head domain of the fibre recognizes a specific surface-exposed cellular receptor (Zhang and Bergelson, 2005). A secondary interaction between the penton base protein and cellular integrins facilitates virus internalization via clathrin dependent endocytosis (Wickham et al., 1993; Li et al., 2001; Soudais et al., 2000; Bangari

and Mittal, 2005). Inside the cell, the virus encounters the acidic environment of the early endosome, prompting capsid destabilization and leading to its partial disassembly (Greber et al., 1993; Meier and Greber, 2003). Partial disassembly frees some components of the capsid, including protein VI that, through its N-terminal helix, destabilizes the endosomal membrane, resulting in the release of viral capsid into the cytoplasm (Wiethoff et al., 2005). The capsid then uses the extensive network of microtubules inside the cell to reach the nuclear periphery, where it docks itself onto the nuclear pore complex, resulting in complete capsid disassembly and release of the viral genetic material inside the nucleus (Trotman et al., 2001; Leopold and Crystal, 2007).

1.5.1 Cellular receptors for adenovirus attachment

As mentioned earlier, the long, trimeric, fibre proteins of adenoviruses mediate high affinity attachments to host cell surface receptors through their distal knob or head domain (Di Guilmi et al., 1995; Henry et al., 1994; Louis et al., 1994; Mei et al., 2002). The shuffling of the fibre domain between adenovirus types is considered as an important event behind the emergence of new adenovirus strains (de Jong et al., 1983; Flomenberg et al., 1987; Kajon and Wadell, 1996).

1.5.1.1 Coxsackievirus and adenovirus receptor (CAR)

CAR is a class I trans-membrane protein with a broad tissue distribution in humans. It has been found on cells of the lung, the pancreas and the central and peripheral nervous systems. However, it is scarce on malignant cells (Tomko et al., 2000). CAR was first identified as an adenovirus receptor for HAdV-2 and HAdV-5 (Bergelson et al., 1997). In deglycosylated form, it has a molecular weight of 38 kDa, while glycosylation confers an additional 8 kDa. It can be divided into three domains: a C-terminal 107-residue intracellular domain, a single trans-membrane helix and an extracellular domain of 216 amino acids. The extracellular region can further be divided into two immunoglobulin-like domains, D1 and D2 (Chretien et al., 1998). The N-terminal D1 domain is necessary and sufficient for fibre binding (Wang and Bergelson, 1999). The attachment of virus to surface-exposed CAR is straightforward, but an interesting twist came when it was found that CAR is expressed on the basolateral surface of polarized epithelial cells, below the tight junction (Cohen et al., 2001). This raises the question on how adenovirus gains access to this location. It was suggested that the virus might gain access through a temporary break in the cell monolayer and subsequently spread by overproducing fibre protein, which disrups CAR-CAR homodimers and thus the tight junctions (Walters et al., 2002).

Except for the members of species B and a few adenoviruses of species D, almost all human adenoviruses can bind to CAR (Cupelli and Stehle, 2011). Besides that, animal adenoviruses infecting chimpanzee and dog also use CAR as their entry receptor (Cohen et al., 2002; Soudais et al., 2000; Tan et al., 2001). The crystal structure of the HAdV-12 fibre knob bound to CAR D1 domain shows the regions of the fibre head domain important for binding (Bewley et al., 1999). They are located on the lateral side of the trimer and include the AB- and DE-loops of one monomer and the FG-loop of a neighbouring monomer. The AB-loop turned out to be the most important, with several patches of conserved amino acids important for CAR-binding, even among divergent fibres of different CAR-binding adenovirus types (Roelvink et al., 1999). The binding of three CAR D1 domains to the trimeric fibre head leads to tighter binding and a low effective dissociation constant (Lortat-Jacob et al., 2001).

The length of the fibre protein as well as its flexibility plays a significant role in CAR attachment. The beta-spiral repeat containing central shaft domain and its juncture with the head domain likely confers flexibility to the overall fibre and it has been observed that decreasing or increasing the number of these repeats in the shaft reduces CAR association (Wu et al., 2003; Shayakhmetov and Lieber, 2000).

1.5.1.2 Sialic acid

Sialic acid is one of the most common terminal modifications of cell surface glycoproteins and gangliosides. It acts as a common target for virus attachment (Matrosovich et al., 2013). Some of human species D adenoviruses use alpha-2,3-linked sialic acid as their attachment receptors (Burmeister et al., 2004). These are serotypes 8, 19 and 37, which are associated with ocular infections. Their fibre heads carry a predicted positive surface charge when compared to other adenoviruses (Arnberg et al., 1997). The fibre head- sialic acid interaction was shown to take place in a charge-dependent manner (Arnberg et al., 2002) and the fibre residues involved in the interaction are located at the top of the trimeric head domain (Burmeister et al., 2004). The weak affinity observed for the interaction (K_d in mM range; Burmeister et al. (2004)) suggests the involvement of multiple sialic acid moieties under actual *in vivo* conditions, as has been suggested for other viruses (Sauter et al., 1989; Dormitzer et al., 2002a).

1.5.1.3 CD46

CD46 is a membrane-linked glycoprotein associated with the complement system of cellular immunity (Liszewski et al., 1996). The extracellular part of CD46 consists of four domains, namely SCR1, SCR2, SCR3 and SCR4 (SCR stands for short consensus repeat;

(Gaggar et al., 2005)). Apart from regulatory functions and binding to C3b/C4b, SCRs also act as receptors for several pathogens such as the measles virus, streptococcal bacteria and herpesvirus (Giannakis et al., 2002; Buchholz et al., 1997; Greenstone et al., 2002; Hsu et al., 1997; Liszewski et al., 2000). A higher level of CD46 expression on malignant cells in humans suggests a strategic adaptation of these cells for progression and survival (Hara et al., 1992; Murray et al., 2000; Gaggar et al., 2005).

CD46 is widely distributed on different nucleated cell types and relatively well accessible on the cell surface. Several species B and one species D adenovirus that doesn't bind CAR have been shown to use CD46 as their cellular receptor (Sirena et al., 2004; Gaggar et al., 2003; Wu et al., 2004a). The crystal structure of human adenovirus 11 (HAdV-11) and 35 (HAdV-35) fibre heads bound to the extracellular region of CD46 reveals the contacting points between two proteins (Persson et al., 2007; Wang et al., 2007). The interaction takes place in a shape complementary fashion, involving SCRs 1 and 2 of CD46 and the DG-, HI- and IJ-loops of the fibre heads.

1.5.1.4 Secondary receptor: integrins

To facilitate virus entry into the cell, adenovirus interacts with vitronectin-binding αvβ3 or αvβ5 integrins (Wickham et al., 1993; Li et al., 2001). Integrins are heterodimeric cell surface glycoproteins composed of α and β subunits, of which there exist 18 and 8 different subunits respectively (Stewart and Nemerow, 2007). Their combinations give rise to several distinct types of integrins, some of which are capable of binding to viral proteins. In the case of adenoviruses, the viral protein component facilitating this interaction is an RGD motif present in the penton base that is conserved in many, if not all, types of adenoviruses (Albinsson and Kidd, 1999). The penton- integrin binding triggers a cascade of intracellular signaling, leading to virus internalization via clathrin-dependent endocytosis (Stewart and Nemerow, 2007). Interestingly, an RGD motif has been found in the fibre proteins of murine adenovirus 1 (MAdV-1) and porcine adenovirus 4 (PAdV-4) (Raman et al., 2009; Kleiboeker, 1995).

1.6 Snake adenovirus 1 (SnAdV-1)

Adenovirus infections have been observed in many reptilian species such as lizards, snakes, turtles and crocodiles; and a range of associated clinical symptoms have been described (Marschang, 2011). SnAdV-1 was isolated from a corn snake (*Elaphe guttata*), which showed clinical signs of pneumonia (Juhasz and Ahne, 1993). Based on the organisation and of its 27.7 kbp genome and phylogenetic considerations, it was grouped under the *Atadenovirus*

genus of the *Adenoviridae* family (Harrach, 2011). SnAdV-1 has a relatively unbiased AT-content, which is in contrast to the previously characterized members of the same genus. However, phylogenetic analysis clearly grouped it with other members of the genus (Farkas et al., 2008). The first Atadenovirus to be fully sequenced was ovine adenovirus (OvAd-D) isolate OAV287 (Vrati et al., 1996). The capsid of OAV287 has some interesting differences when compared to that of other adenovirus genera (Löser, 1999; Pantelic et al., 2008).

SnAdV-1 has one fibre gene, encoding a predicted polypeptide of 415 amino acid residues (Farkas et al., 2008). However, in this doctoral work, the fibre gene product was found to be only 345 amino acids long (Singh et al., 2013b). Its primary sequence doesn't have much similarity with fibres of other adenoviruses; particularly the C-terminal part is highly divergent (only between 12 and 18% sequence identity to adenovirus fibre heads of known structure). Despite the low sequence identity, some likely features can be identified. One is the presence of a short stretch of hydrophobic residues (13-PVYPFG-18) in the N-terminal tail region of the protein, which may be involved in fibre-penton interactions. A conserved cluster of basic amino acids was shown to be present in human adenoviruses (Chroboczek J, 1995) and their role as a nuclear localization signal was demonstrated (Wu et al., 2004b). Similarly, a short stretch of basic residues (3-KIKR-6) is apparent very near the N-terminus of the SnAdV-1 fibre. Potential beta-spiral repeats are present between residues 38 and 224. Secondary structure predictions suggested that a putative loop region between residues 226 and 236 containing several prolines might separate the shaft from the head domain. The putative head domain would then be 111 residues long and shorter than all adenovirus fibre heads with known structures. The OvAd-D fibre head, which belongs to the same genus as of SnAdV-1, is also significantly smaller than other adenovirus fibre heads, as shown by electron microscopy (Pantelic et al., 2008).

1.7 Turkey adenovirus 3 (TAdV-3)

Turkey adenovirus type 3 (TAdV-3) belongs to the *Siadenovirus* genus of the *Adenoviridae* family. It is also called turkey hemorrhagic enteritis virus (THEV) due to the clinical symptoms associated with its infection (Harrach, 2011). Electron microscopy of purified virions revealed adenovirus-like virions (Itakura and Carlson, 1975; Carlson et al., 1974; Tolin and Domermuth, 1975). TAdV-3 was shown to have one fibre (van den Hurk, 1992). Serologically indistinguishable strains of TAdV-3 are known to have variable levels of pathogenicity. Virulent forms cause hemorrhagic enteritis in turkeys and marble spleen disease in pheasants (Domermuth et al., 1979; Palya et al., 2007), while avirulent or less virulent

forms don't and can be used as vaccines (Domermuth et al., 1977; Silim et al., 1978; Beach et al., 2009b).

A comparison of the genome sequences of virulent and avirulent TAdV-3 strains has revealed mutations in the fibre sequence (Beach et al., 2009a). The 454-residue fibre protein of TAdV-3 contains a predicted tail domain (residues 1-45), which is likely to interact with the penton base, and a predicted shaft domain with putative triple beta-spiral repeats between residues 46 and 303. The remaining part (comprising amino acids from 304 to 454) likely forms the head domain. Two of the aforementioned mutations in the fibre sequence are located in the predicted head domain, namely residues 354 and 376, which are methionines in the avirulent form and isoleucine and threonine in a virulent strain, respectively.

1.8 Lizard adenovirus type 2 (LAdV-2)

Isolation and laboratory propagation of reptilian adenoviruses is challenging, obstructing their characterisation (Ogawa et al., 1992; Juhasz and Ahne, 1993). Adenoviruses have been described in several lizard species (Ogawa et al., 1992; Julian and Durham, 1982; Wellehan et al., 2004), but sequence information has been limited to the DNA polymerase gene (Wellehan et al., 2004). Recently, Papp et al. have reported isolation of lizard adenovirus-1 from Gila monsters (*Heloderma suspectum*) and lizard adenovirus-2 (LAdV-2) from a Mexican beaded lizard (*H. horridum*) (Papp et al., 2009).

Recently we have obtained genomic and proteomic information for LAdV-2 through personal communication with the groups of Dr. Balazs Harrach (VMRI, Hungary) and Dr. Carmen San Martin (CNB-CSIC, Madrid). The 32 kbp genome of LAdV-2 revealed an interesting feature; that of the presence of two fibre genes coding for two distinct fibre proteins (Penzes et al., 2014). Analysis of purified virions by mass spectrometry and electron microscopy evidenced the presence of a short fibre on most vertices, and three long fibres on the remaining vertices. LAdV-2 is the first characterised Atadenovirus so far that has two distinct fibres. The fibre-1 gene of LAdV-2 codes for a polypeptide of 331 amino acids with a monomeric weight of 34.7 kDa, whereas the predicted length for fibre-2 gene product is 433 amino acids, with a molecular weight of 45.7 kDa. The two proteins differ significantly with each other and have a sequence identity of just 20%.

Sequence analysis and secondary structure predications for both fibre protein shows characteristics of typical adenovirus fibre proteins with the presence of triple beta-spiral repeats in a central shaft domain (van Raaij et al., 1999b). In the case of fibre-1 the predicted tail domain ranges from 1-40 while the putative shaft is from amino acid 45 to 210. The fibre-

2 is longer than fibre-1 by 102 amino acids and its putative tail domain encompasses residues 1-45. The shaft domain is estimated to encompass residues 46-260. The remaining C-terminal portions of the respective proteins will probably form the head domains. Instances of two fibre proteins in adenoviruses have only been described in fowl adenovirus type-1 (FAdV-1), a member of *Aviadenovirus* genus, and in members of species F human mastadenoviruses (Hess et al., 1995; Kidd et al., 1993).

1.9 Murine adenovirus type 2 (MAdV-2)

The first report of the presence of adenovirus-like viral particles in mice was documented by Hartley and Rowe (Hartley and Rowe, 1960). Experimentally, the virus, called MAdV-1, was shown to cause encephalomyelitis in newborn or immunodeficient mice, as well as in several strains of adult mice (Guida et al., 1995; Wigand, 1980). Following MAdV-1, murine adenovirus types 2 and 3 (MAdV-2, MAdV-3) were identified. All three viruses are mastadenoviruses (Klempa et al., 2009; Sugiyama et al., 1967; Hashimoto et al., 1966). While MAdV-1 and MAdV-2 were isolated from house mice (*Mus musculus*), MAdV-3 was isolated from a striped field mouse (*Apodemus agrarius*).

MAdV-2 was isolated from the feces of inbred, healthy house mice and *in vitro* studies showed its ability to cause cytopathic changes in mouse cell lines, particularly that of gastrointestinal origin, but not in monkey or human cell lines (Hashimoto et al., 1966; Takeuchi and Hashimoto, 1976). The tissue specificity of MAdV-2 is different from MAdV-1, which targets endothelial cells of the brain and spinal cord and cells of mononuclear phagocytic lineage (Lenaerts et al., 2009). MAdV-2 growth in cell cultures was found to be highly inefficient, thus limiting its availability for experimental purposes (Hamelin et al., 1988). *In vivo* infection of MAdV-2 in healthy mice does not appear to cause disease, at most some diarrhea (Sugiyama et al., 1967).

The genome sequence of MAdV-2 has been determined (Hemmi et al., 2011), which confirmed MAdV-2 is a Mastadenovirus and also showed that the genome of MAdV-2 at 35,203 bp is the largest of the three sequenced murine adenovirus types (Hemmi et al., 2011; Meissner et al., 1997; Klempa et al., 2009). All major structural genes are conserved, as are the mastadenovirus-specific proteins IX and V. Interestingly the RGD motif is absent from the penton base. MAdV-2 has one fibre gene, which codes for a 787-residue long polypeptide. The fibre of MAdV-2 has more amino acids than those of MAdV-1, MAdV-3, HAdV-2, HAdV-3 and HAdV-5. Secondary structure prediction of fibre suggests a short tail domain (residues 1-50), a long central shaft domain (residues 53-630) and a C-terminal head domain

(residues 633-787). Between residues 40 and 47, a short proline-rich region may be involved in helping the fibre attach to the penton base. The C-terminal domain doesn't share significant sequence homology with other known fibre head structures (the sequence identity ranges between 10-16%). As mouse adenoviruse could be interesting model system to study adenovirus pathogenicity and given the crucial role of fibres in mediating adenovirus-host interactions, structural characterisation of MAdV-2 fibre may advance our knowledge about specific virus-host interaction.

Objectives

2. Objectives

The aim of this thesis work was to structurally and functionally characterise the outer capsid proteins of several animal adenoviruses; snake adenovirus 1, turkey adenovirus 3, lizard adenovirus 2 and murine adenovirus 2. The structures of these proteins might give clues about receptor-binding and better understanding of adenovirus biology. In the future, this information may also help the design of adenovirus-based vectors for gene and cancer therapies. The following specific objectives were pursued:

1. Cloning, expression, purification and crystallization of the snake adenovirus 1 fibre protein. Structure solution of the receptor binding head domain. Comparison of the structure with known structures. Identification of potential ligands
2. Expression, purification and crystallization of avirulent and virulent turkey adenovirus 3 fibre proteins. Structure solution of the receptor binding head domains. Comparison of the structures with each other and with known structures. Identification of potential ligands, determination of their affinities for the fibre head domain and structure solution of fibre head-receptor complexes.
3. Expression, purification and crystallization of the lizard adenovirus 2 long and short fibre proteins. Structure solution of the head domains of the short fibre. Comparison of the structure with the snake adenovirus 1 fibre head.
4. Expression, purification and crystallization of the murine adenovirus 2 fibre protein. Structure solution of the C-terminal domain. Comparison of the structure with known adenovirus fibre head structures. Identification of potential ligands.
5. Cloning, expression, purification and crystallization of the snake adenovirus 1 LH3 protein.

Materials and Methods

3. Materials and Methods

3.1 Biological and non-biological materials

3.1.1 Bacterial cells

For cloning and protein expression purpose the following bacterial strains were used during the course of this work:

Table 2: *Escherichia coli* strains used for plasmid amplification and protein expression

E. coli strain	Genotype	Description	Supplier
TOP10	F^- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697galU galK rpsL (StrR) endA1 nupG	*E. coli* B strain derivative; used for cloning and plasmid propagation.	Invitrogen (Madrid, Spain)
BL21(DE3)	F^- ompT hsdSB (rB^-mB^-) gal dcm (DE3)	*E. coli* B strain derivative; used for protein expression.	Novagen (Merck, Darmstadt, Germany)
Rosetta(DE3)pLysS	F^-ompT $hsdS_B(r_B^- m_B^-)$ gal dcm (DE3) pLysSRARE (Cam^R)	*E. coli* B strain derivative; used for protein expression.	Novagen (Merck, Darmstadt, Germany)
B834(DE3)	F^-ompT $hsdS_B(r_B^- m_B^-)$ gal dcm met (DE3)	*E. coli* B strain derivative; Methionine auxotrophs; used for protein expression.	Novagen (Merck, Darmstadt, Germany)
C41(DE3)	F^- ompT hsdSB ($rB^- mB^-$) gal dcm (DE3)	*E. coli* BL21(DE3) derivative; used for protein expression.	Lucigen (Middleton, WI, USA)
JM109(DE3)	endA1, recA1, gyrA96, thi, hsdR17 (r_k ,m_k^+), relA1, supE44, λ–, Δ(lac-proAB), [F´, traD36, proAB, $lacI^qZ\Delta M15$], lDE3	Used for protein expression.	Promega (Madison, WI, USA)
XL10-Gold	endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet^R F'[proAB $lacI^qZ\Delta M15$ Tn10(Tet^R Amy Cm^R)]	Used for protein expression.	Stratagene (Agilent Technologies, Santa Clara, CA, USA)

The *E. coli* TOP10 strain was used for cloning and plasmid propagation as it carries the *hsdR* gene that enhances the efficiency of transformation of unmethylated PCR amplified products and the *recA1* (a mutant of *recA*) gene to reduce chances of getting unwanted recombination in the cloned genes respectively.

E. coli strains that contain a lysogen of λDE3 were used for protein expression in an inducer-controlled manner. The λDE3 prophage contains a gene for the T7 RNA polymerase under control of the strong lacUV5 promoter. In non-induced conditions, the lac repressor protein, coded by the *lac*I gene present in the pET-type expression vectors used in this thesis as well as in the host chromosome, binds to the lac operator site situated upstream of the gene of interest, repressing transcription by T7 RNA polymerase (Bell and Lewis, 2000). Lactose and its more stable analogue IPTG bind to the lac repressor protein, leaving it incapable to bind to the operator site. The T7 promotor sequence is now readily accessible to the T7 RNA polymerase, allowing production of the target mRNA and expression of the target protein.

Sometimes expressing recombinant proteins becomes limited due to the fact that certain tRNAs are scarce in *E. coli*. Rosetta(DE3)pLysS was used to supply extra amounts of these tRNAs in order to obtain maximum expression in case the gene encoding the target protein posses codons which are rare in the bacterial host. The plasmid pLysS encodes T7 bacteriophage lysozyme, which attenuates T7 RNA polymerase activity under non-induced conditions, hence lowering the basal expression of target proteins that may be toxic to *E. coli*.

3.1.2 Protein expression vectors

pET28a(+) and pET28c(+) (Novagen, Merck, Darmstadt, Germany): vectors used for expressing target proteins in bacterial cells. These pBR322-derived vectors were used to express target proteins with an N-terminal 6xHis and T7 tag (M GSSHHHHHHS SGLVP RGSHM ASMTG GQQMG RGS). The two vectors only differ by a frame-shift in their multiple cloning site region. The pET system expression vectors use the T7 RNA polymerase provided by the host cells to express the target protein in a very efficient manner. Leaky, unwanted expression prior to induction is controlled by inclusion of the *lac*I coding segment in the vector. Upon full induction using IPTG, most of the bacterial resources are directed to express the target protein and within a few hours the target protein can comprise more than 50% of the bacterial total protein content. The kanamycin resistance gene present in the pET28a(+) and pET28c(+) vectors provides an easy way to select for positive clones on kanamycin containing plates and continuing selection of plasmid presence in liquid growth media.

3.1.3 Bacterial growth/culture media

a) **LB medium**: 10 g/l tryptone, 5 g/l yeast extract, 10 g/l sodium chloride; used for growing bacterial cells. Kanamycin was supplied at a final concentration of 50 μg/ml to prevent growth of non-specific bacteria.

b) **LB-agar**: 10 g/l tryptone, 5 g/l yeast extract, 10 g/l sodium chloride, 15 g/l bacteriological agar; LB media supplemented with agar was used to grow bacterial colonies on a solid surface in order to select single colonies. Antibiotic was added as above to select for plasmid-containing colonies.

c) Transformation and storage solution (TSS): LB with 12% (w/v) PEG 8000, 5% (v/v) DMSO, 25 mM magnesium chloride and 1% (v/v) glycerol; this solution was used to prepare and store chemically competent bacterial cells for plasmid transformation.

3.1.4 Buffers

The following general-purpose buffers were utilized during the course of this work:

a) **SDS-PAGE electrophoresis buffer**: 25 mM Tris-HCl pH 8.3, 192 mM glycine, 0.1% SDS.

b) **TAE buffer (1X)**: 40 mM Tris-HCl pH 8.3, 1.1% acetic acid and 0.001% EDTA

c) **DNA loading buffer (6X)**: 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, 30% (v/v) glycerol

d) **Protein sample loading buffer (1X)**: 60 mM Tris-HCl pH 6.8, 2% SDS, 2.5% beta-mercaptoethanol, 0.025% bromophenol blue, 10% glycerol

3.2 Methods

3.2.1 Preparation of chemically competent *E. coli* cells and performing transformation

To obtain transformation-competent *E. coli* cells, 10 ml of LB medium were inoculated with bacteria taken from a glycerol stock and allowed to grow overnight at 37 °C. The following day 100 ml fresh LB medium was inoculated with the overnight culture in a ratio of 1:1000 and allowed to grow until the O.D. reached 0.3 measured at 600 nm. The cells were harvested by centrifuging at 3000 g for 10 min at 4 °C and resuspended in 10 ml TSS buffer. Aliquots of 100 μl each were prepared and stored at -80 °C.

Transformations were carried out using the heat shock method: 100 μl of competent cells were mixed with 50-75 ng of plasmid DNA and the mixture was put on ice for 30 minutes.

The mixture was heated for 60 seconds at 42 °C, followed by quick cooling on ice for 2-3 minutes. LB medium (500 µl) was then added and the mix was incubated for 1 hour at 37 °C with shaking. Finally the mix was centrifuged, around 400 µl of the supernatant was removed and the cells were plated on pre-warmed kanamycin-containing LB-agar plates.

3.2.2 Working with DNA

3.2.2.1 DNA amplification by PCR

Gene amplification reactions were carried out by a three-step polymerase chain reaction in a MJ mini personal thermo cycler (Bio-Rad, Hercules, CA, USA) using reagents supplied in the *Taq* DNA polymerase kit (New England Biolabs, Ipswich, Massachusetts, USA). These reagents include *Taq* DNA polymerase, reaction buffer, deoxynucleotides and ready-to-use DNA marker. Each reaction consisted of the following steps:

3-step cycle:	
Initial denaturation: 94 °C, 5 min **Denturation**: 94 °C, 30 sec ***Primer annealing**: (T_m-5) °C, 1 min **Extension**: 72 °C, 1min for each kb of expected fragment	30 cycles
Final extension: 72 °C, 10 min	

* The primer annealing temperature depends on the calculated melting temperature (Tm) of the oligonucleotides used as given by: T_a Optimal = 0.3 x(T_m of primer) + 0.7 x(T_m of product) – 25. Generally an annealing temperature 5 °C below lowest Tm of the pair of primer is used

Amplified products were subjected to 1 % (w/v) agarose gel electrophoresis and purified using a GeneJET gel extraction kit (Thermo Fisher, Pittsburgh, USA).

3.2.2.2 Plasmid purification

Plasmids were purified from 10 ml bacterial cultures grown overnight with the alkaline lysis method, using a Wizard Plus SV Miniprep (Promega, Madison, WI, USA) or GeneJET Plasmid Miniprep (Thermo Fisher, Pittsburgh, USA) kit and following the supplied protocol.

3.2.2.3 Restriction digestion and ligation

Amplified gene products and circular plasmids were subjected to restriction digestion following the enzyme manufacturers guidelines. Digested products were gel purified using a GeneJET gel extraction kit (Thermo Fisher, Pittsburgh, USA).

Ligation reactions were carried out either at room temperature for 3 hours or at 16 °C overnight using T4 DNA ligase and the supplied ATP-containing ligase buffer (Promega, Madison, WI, USA). An insert to vector ratio of 3:1 was maintained in which 100 ng of linearised vector was used and the required amount of digested insert was calculated using the following formulae:

$$ng\ of\ insert = \left(\frac{ng\ of\ vector \times kb\ of\ insert}{kb\ of\ vector} \right) \times insert\ to\ vector\ ratio$$

3.2.2.4 Site-directed mutagenesis

Site-directed mutagenesis experiments were carried out using a MJ Mini thermo cycler PCR (Bio-Rad, Hercules, CA, USA) in a total reaction volume of 50 µl comprising of 100 ng of recombinant plasmid and *Pfu* Turbo DNA polymerase (Agilent Technologies, Santa Clara, California, USA). Specific methodologies adopted for each case are given in respective sections.

3.2.3 Protein methods

3.2.3.1 SDS-PAGE electrophoresis

The electrophoretic mobility of biological macromolecules in polyacrylamide gels depends on their charge and size. However, in SDS-PAGE, protein molecules attain a uniform negative charge due to the presence of SDS and their mobility only depends on their size.

A polyacrylamide concentration of between 10 and 15% was used to prepare the resolving part of the gel, depending on the size of the target protein. After the run, gels were stained by 30 minutes of gentle shaking in a solution containing 0.25% Coomassie Brilliant Blue, 33% methanol and 10% glacial acetic acid. Destaining was for variable times using the same solution but without Coomassie Brilliant Blue. A molecular weight marker mixture (PageRuler unstained low range or broad range protein ladder; Thermo Fisher, Pittsburgh, USA) was run simultaneously to estimate the molecular weight of the visualized proteins.

3.2.3.2 Protein expression, cell harvesting and disruption

Expression of proteins was carried out in various *E. coli* strains using the following scheme (unless otherwise mentioned): 10 ml of starter culture was prepared from transformed cells and allowed to grow overnight with shaking at 37 ºC. Four 0.7 litre cultures in LB medium were inoculated in a ratio of 1:1000 and allowed to grow aerobically at 37 °C to an O.D. of 0.6–0.8 measured at 600 nm. The temperature was reduced by cooling on ice and the cultures were induced with 1 mM IPTG followed by overnight growth at 16 °C. Cells were harvested by centrifugation at 5000 x *g* for 15 min and resuspended in lysis buffer consisting of 50 mM Tris–HCl pH 7.5, 500 mM sodium chloride, 20 mM imidazole, 10% (v/v) glycerol. Lysis was performed using a French press cell disruptor (three passes). Insoluble material was removed by centrifugation (20000 x *g*, 45 min) and supernatant was collected for further purification.

3.2.3.3 Protein purification

A two-stage purification strategy was adopted in order to obtain crystallization-grade protein for all the proteins studied in this thesis. Histidine-tagged proteins were first purified by nickel-nitrilotriacetic acid (Ni-NTA) affinity column chromatography followed by ion exchange chromatography (IEC). The six-histidine tag is very small and at a working pH of 7.5-8.0 is usually neutrally charged, not affecting protein folding or compartmentalization.

a) **Ni-NTA metal affinity chromatography:** Nitrilotriacetic acid is a tetradentate chelating agent which occupies four of the six ligand binding sites in the coordination sphere of the nickel ion (Figure 6). Thus two sites are free to interact and bind histidines presented by the target protein. On one side NTA is made bound to agarose bead base while the other side is exposed. Once the tagged protein is bound, the column is washed to eliminate any non-specifically bound protein. The histidine analog imidazole can be used to competitively elute bound six-histidine tags, thus eluting the target protein. The following protocol was adopted to purify target proteins: 1.5 ml of Ni-NTA resin (Jena Bioscience GmbH, Jena, Germany) was added to the supernatant collected after cell disruption and the sample was left for incubation with gentle shaking for 1 h. The suspension was poured into an empty column and washed twice with 10 column volumes of buffer consisting of 50 mM Tris–HCl pH 7.5, 500 mM sodium chloride, 30 mM imidazole, 10% (v/v) glycerol. This was followed by a step gradient elution with imidazole (50, 100, 200, 300, 400, 500 and 1000 mM) in the same buffer (each step was

with 2 column volumes of the respective elution buffer). The fractions containing the target protein were pooled and dialysed, using 10 kDa cutoff dialysis tubing (Sigma-Aldrich, St Louis, Missouri, USA). The choice of dialysis buffers was based on protein's predicted pI and is mentioned in respective sections.

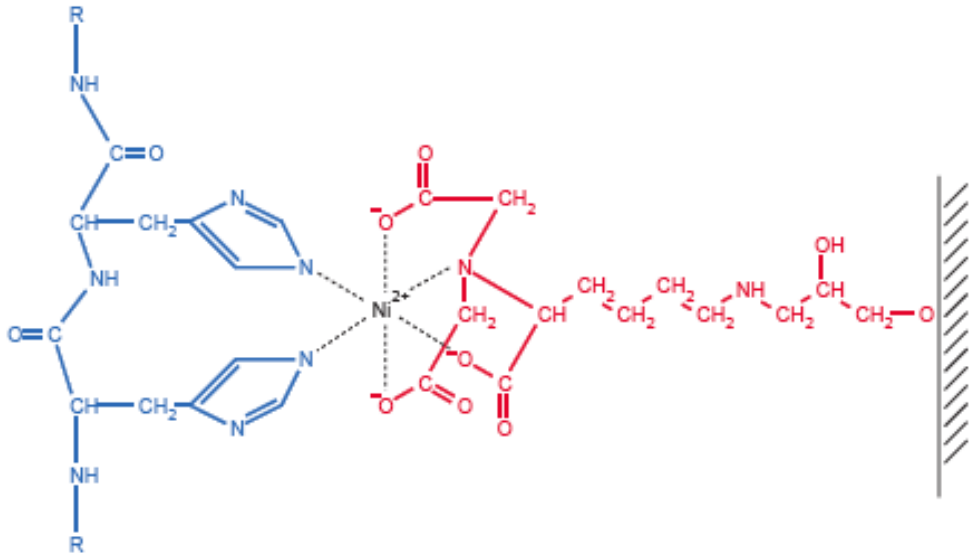


Figure 6: Ni-NTA chromatography. Nickel ion interaction with histidine residue (in blue). (from Qiagen Ni-NTA manual)

b) **Ion exchange chromatography (IEC):** Proteins, being amphoteric in nature, contain both acidic and basic functional groups. The local and overall charges of a protein molecule depend on the pH of their environment. Ion exchange chromatography exploits this behaviour and can be classified broadly into two types: anion exchange chromatography and cation exchange chromatography. In anion exchange, negatively charged protein patches are attracted towards a positively charged solid support while in cation exchange a negatively charged solid support attracts a positively charged molecule (Figure 7). Ion exchange chromatography was carried out on a ÄKTA purifier FPLC (GE Healthcare Bio-Sciences AB, Uppsala, Sweden), using Resource Q and Resource S columns (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) for anion and cation exchange chromatography, respectively. Protein samples were dialyzed against suitable buffers (mentioned in respective sections) and linear gradient elutions with 0-1 M sodium chloride was performed in the same buffer. Collected peak fractions were desalted and concentrated using Amicon Ultra concentrators (Millipore Iberica, Madrid, Spain), applying three washes with the buffer of choice and stored at 4 $^{\circ}$C prior to crystallization trials.

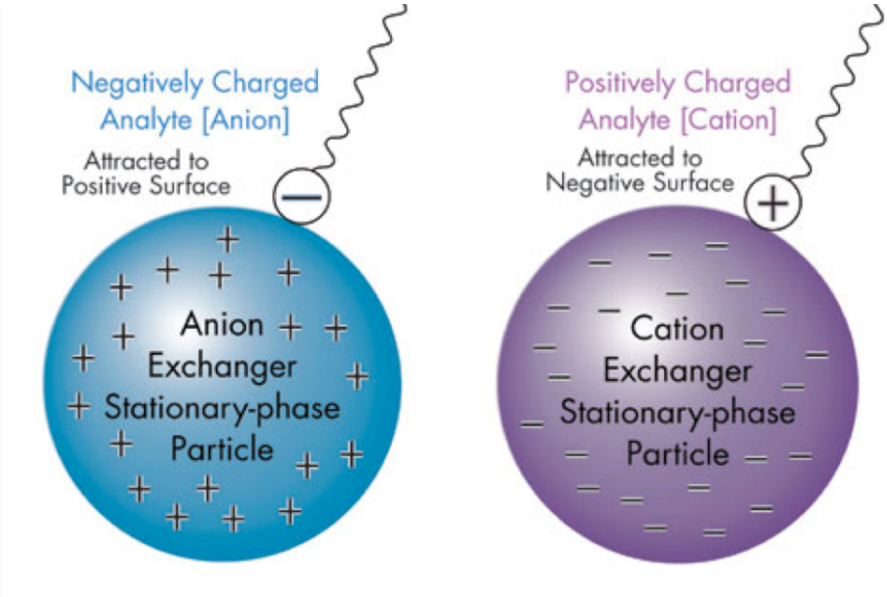


Figure 7: Anion and cation exchange chromatography

3.2.3.4 Neoglycoconjugate (NGC) microarrays

It has been shown that adenovirus may use cell surface glycans for their entry into cell via fibre mediated interactions (Arnberg et al., 2000). With the aim of identifying potential glycan ligands for the fibre head proteins under study, neoglycoconjugate (NGC) microarray experiments were carried out by Dr. Michelle Kilcoyne in the group of Dr. Lokesh Joshi at NUI, Galway, Ireland. The microarray slides used in the experiments presented different simple to complex glycans attached on protein or lipid backbones through linkers or spacers. Proteins were titrated in TBS-T buffer (1 mM $CaCl_2$, 1 mM $MgCl_2$, 0.05% Tween 20 pH 7.2) to determine optimal concentration and later 10 μg/ml was found suitable for experiments as at higher concentrations significant background noise was detected. In a similar way, the concentration of the probing anti-6X His IgG-CF640R (Sigma–Aldrich, St Louis, Missouri, USA) antibody was optimized.

All experiments were carried out in triplicate, each comprising of two steps: slides were incubated in TBS-T for 1 hour at 23 °C followed by three washes in TBS-T for 2 minutes and an onetime wash with TBS. After the wash slides were allowed to dry and a second incubation with monoclonal anti-6X His IgG-CF640R was performed for 1 hour at 23 °C. The antibody detects the 6xHis tag present on the proteins. Binding inhibition assays were performed in the presence of suitable glycans.

3.2.3.5 Saturation transfer difference NMR spectroscopy

The saturation transfer difference experiment is based on the transfer of magnetization between the protein, which is selectively irradiated by a train of radiofrequency pulses, and any binding ligands (Meyer and Peters, 2003). For complex ligands, such as oligosaccharides, those atoms lying closer in space to the protein surface receive a higher amount of magnetization, providing information on the ligand binding epitope (Mayer and Meyer, 1999).

STD-NMR experiments with fibre head proteins were conducted at the Centro de Investigacion Biologia (CIB-CSIC) Madrid, by Alvaro Berbis in the group of Dr. Jesús Barbero. NMR spectra were acquired at 298 K in a Bruker AVANCE 500 MHz spectrometer equipped with a 5 mm inverse probe head. Experiments were performed in fully deuterated 10 mM potassium phosphate buffer, pH 7.7. Samples contained 1.5 mM of ligand and 15 μM of fibre head protein. Data were analyzed with Bruker TopSpin 3.0 and figures were made with MestReNova v.8.0.2.

3.2.3.6 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry experiments were carried out at IQFR-CSIC, Madrid, Spain by Dr. Margarita Menendez. Titrations were performed at 25 °C with a Microcal VP-ITC microcalorimeter (GE Healthcare Bio-Sciences AB, Uppsala, Sweden). The proteins were exhaustively dialyzed before against 10 mM MES, pH 6.0 (in case of TAdV-3 fibre) and 10 mM Tris-HCl, pH 8.0 (in case of MAdV-2 fibre), and sugar solutions were prepared with the final dialysates. Titrations were performed by stepwise injections of the ligand-containing solution (21-24 mM) into the reaction cell loaded with the protein solution at concentrations of 78–474 μM. When required, a second set of injections was coupled to the first one after refilling the injection pipette with the same sugar solution, and ConCat32 program was used to couple the raw data of both series. The heat of ligand dilution was determined separately and subtracted from the total heat produced following each injection. Protein concentrations were determined spectrophotometrically at 280 nm using the calculated molar extinction coefficient of monomer. Titration data were analyzed with the Microcal-ITC ORIGIN software.

3.2.4 Crystallographic methods

3.2.4.1 Protein crystallization

Crystallization of protein molecules can't be put under the same umbrella with that of small molecules as both differ significantly in their physical and chemical properties. In contrast to stable and rigid small molecules, protein molecules are highly labile, dynamic and show more complexity in the way they interact with each other. Their crystals, as well, demonstrate differences in internal and external properties. Small molecule crystals are physically stronger, more brittle and they usually diffract X-rays much better than protein crystals. Such properties can be explained by their highly ordered lattice system and strong crystal contacts. On the other hand, the total number of different bonding interactions within a protein crystal is far less than in a small molecule crystal (McPherson, 2008), resulting in an overall weak lattice force and thus explaining their softness, fragility and relatively poor diffraction properties. Moreover, protein crystals are also very sensitive towards changes in external conditions such as pH, temperature and osmolarity, owing to their high solvent content (Littlechild, 1991) usually between 30 and 80% (Matthews, 1985)

A primary consideration in protein crystallization is sample quality. A sample should satisfy properties that are critical for successful crystallization: homogeneity, stability and solubility. Availability of identical units of target molecule in the sample is a must, as it will form the periodic lattice of the crystal. The presence of any impurity in the sample may hamper this process. These impurities could be externally or internally originated, internally if the molecules, independent of one another, change their state by some means. This will eventually lead the solution from a state of homogeneity to heterogeneity. Therefore the molecule should be maintained in a stable state. A molecule should also be highly soluble and in a monodisperse solution before starting crystallization trials. There should not be any sort of aggregates, molecular clusters or non-specific oligmers present in the solution (Bergfors, 1999).

Although small molecules and protein macromolecules are very different in their physiochemical properties, both follows same basic rule for crystallization. Their solutions must be in a state of supersaturation in order to initiate crystal formation. The solution can be forced to attain this state by a variety of means such as altering the temperature, the pH and the precipitant concentration of the solution. This would eventually lead to a change in protein solubility, pushing the system towards supersaturation.

Vapour diffusion is an approach where changing the precipitant and protein concentration simultaneously may lead to supersaturation. I employed the sitting drop method, where the drop is in a sitting position (Figure 8). In both cases a concentration gradient develops between the drop containing the protein and the mother liquor in reservoir as the concentration of precipitant is higher in the reservoir when compared to the drop. This will lead water to evaporate from the drop thus lowering its volume and increasing protein and precipitant concentration. At some point, the protein concentration may attain a state of supersaturation and if all other parameters, which are important for crystallization, such as pH, temperature, additives etc., are optimal, the protein may crystallize.

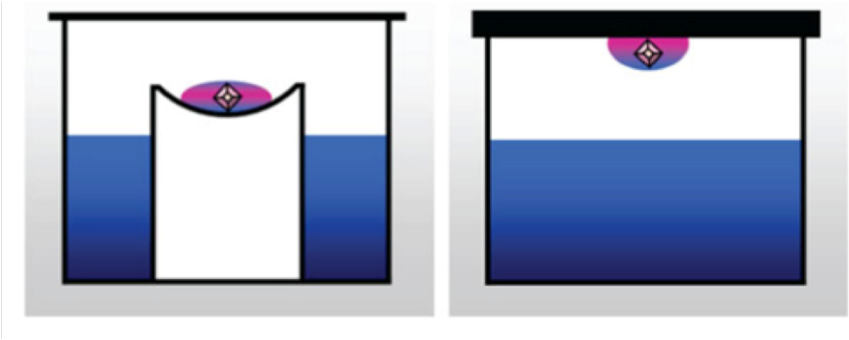


Figure 8: Sitting drop and hanging drop techniques for crystallization of biological macromolecules. The drop containing the macromolecule is shown in pink whereas the mother liquor containing reservoir is in blue.

3.2.4.2 Crystal harvesting

Depending on their growth conditions, crystals were either cryoprotected or harvested directly for data collection as mentioned in respective sections. Crystals were mounted in LithoLoops (Molecular Dimensions, Newmarket, Suffolk, England) or MicroMounts (Mitegen, Ithaca, New York, USA) and vitrified at 100 K. X-ray diffraction data were collected at 100 K.

3.2.4.3 X-ray diffraction data collection

For crystals obtained during the course of this thesis work, data sets were collected at different synchrotron beamlines:

A. ***European Synchrotron Radiation Facility (ESRF), Grenoble, France****:* Beamlines BM14, BM30A, ID14-EH1, ID14-EH4 and ID29 were accessed for data collection.

B. ***ALBA synchrotron, Barcelona, Spain****:* The beamline dedicated for macromolecular crystallography, BL13-XALOC, was used.

C. ***Soleil synchrotron, Paris, France****:* Beamline Proxima-1 was employed.

In addition, the home X-ray source at the CSIC Institute for Physical Chemistry (IQFR-CSIC, Madrid) was used in some cases for crystal diffraction testing.

3.2.4.4 X-ray diffraction data processing

Diffraction data sets were processed with different computational programs. In order to retrieve structural information out of the diffraction spots some basic pre-processing is required which are:

1. *Data indexing*: The main purpose behind this step is to determine the cell parameters of the unit cell present in the crystal by the position of reflection peaks. Simultaneously the crystal symmetry (i.e. space group) can also be inferred. Usually a few images are sufficient for assigning Miller indices, although sometimes the initial space group assignment must be revisited in light of the complete dataset. MOSFLM (Kabsch, 2010; Battye et al., 2011) program was used to index data sets collected during this thesis work unless otherwise mentioned.
2. *Data integration*: Once the space group and unit cell parameters are determined all recorded each measured spot is assigned an intensity value. In this thesis work, data integration was carried out using MOSFLM (Kabsch, 2010; Battye et al., 2011) unless otherwise mentioned.
3. *Merging and scaling*: Two or more diffraction images may contain symmetrically equivalent reflection peaks as the crystal is being rotated to collect redundant dataset. These equivalent peaks are identified and relative images are assigned a uniform intensity scale. Finally these intensities are converted into structure factor amplitudes; achieving structure factor from integrated reflections is called data reduction and now the single output file is ready to be used for initial phasing process. I used POINTLESS, SCALA and TRUNCATE (Evans, 2011), all of which are part of CCP4 (Winn et al., 2011), for data reduction during this work.

   Additionally, SCALEIT (Howell and Smith, 1992; Winn et al., 2011) was used to compare native data with derivative data, where applicable.

3.2.4.5 Structure determination

The ultimate goal in macromolecular crystallography is to obtain structural information, which is defined in terms of electron densities and that we interpret as atomic positions in

three-dimensional space. What are recorded in a diffraction experiment are intensities of the diffraction spots originating from the crystal unit cell. Once having the intensities, I(*hkl*), we can calculate the structure factor amplitudes of the reflections, F(*hkl*), by the relation:

$$F(hkl) = \sqrt{I(hkl)}$$

The amplitudes can be related to the electron density ρ(*xyz*) in the crystal by a Fourier transformation:

$$\rho(x\ y\ z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} \left| F(h\ k\ l) \right| \exp[-2\pi i\,(h\cdot x + k\cdot y + l\cdot c) + i\alpha(h\ k\ l)]$$

To resolve the above equation the missing parameters α(*hkl*), or the phase angles of each structure factor, must also be determined. However, there is no way to measure α, so in order to calculate it additional methods had to be developed:

1. ***MIR/SIR:*** There are methods by which phase can be determined by substituting one or several atoms in the native protein crystal with heavy atoms. These derivative crystals can be prepared by soaking native crystals with suitable heavy atoms or by co-crystallization. In single or multiple isomorphous replacement techniques (SIR/MIR) interpretation of difference Patterson map reveals the location of introduced heavy atom in the unit cell of the derivative crystal. Having the structure factors for both heavy atom derivative crystal (F_{ph}) and the heavy atom (F_h) alone as well as the phase for the heavy atom alone (α_H), the phase for the protein can be calculated by employing the following relation:

$$\alpha_P = \alpha_H + \cos^{-1}\left[\left(F_{PH}{}^2 - F_P{}^2 - F_H{}^2\right) \div 2F_P F_H\right]$$

The absolute requirement for the above method to work is that both the native and derivative crystals should be isomorphous. This means that their space groups and unit cell dimensions as well as the positions of the constituting atoms, except for the introduced heavy atoms, should be same.

2. ***Molecular Replacement:*** This method requires a structural model homologous to the target protein molecule. The possible solution space is then searched using the model

applying methods first described by Michael Rossmann and David Blow (Rossmann, 1962). Statistically, a sequence based similarity greater than 25% is required to apply this method successfully; a good degree of structural similarity between the search model and the target protein is also essential for this method to work successfully.

3. ***Anomalous dispersion methods:*** Phase can also be calculated by using anomalous scattering property of heavy atoms introduced in protein crystal. Electron dense heavy atoms like Hg, Se, Pb etc. scatter the X-ray beam differently than usual atoms present in protein molecules. Thus the intensity difference between Bijvoet pairs: $|F_H(+)|^2$ and $|F_H(-)|^2$ can be exploited to determine the unknown protein phase. In order to do so, it is obvious that the intensity difference between Bijvoet pairs should be as high as possible and therefore the wavelength is selected very carefully keeping in mind the absorption edge of incorporated heavy atoms. Data sets can also be collected at one or more wavelengths. In the case of single wavelength data collection the technique is called single-wavelength anomalous dispersion (SAD) phasing whereas the multi-wavelength anomalous phasing (MAD) term is used if more than one datasets were used for structure solution (Hendrickson, 1985). Synchrotron beamlines, where the X-ray can be tuned, are normally necessary to perform a SAD or MAD experiment. One advantage of this technique over SIR or MIR is that isomorphism between native and derivative crystals is not essential.
4. ***Direct methods:*** These methods can only be applied where there is some prior knowledge of phases for some reflections are available. For protein structure solution the requirement is of very high-resolution data approaching to 1.2 Å or better. This implicates restriction of their application to very small molecules, as in general, resolutions decreases with increasing complexity.

Besides these methods there are other methods such as SIRAS and MIRAS where isomorphous replacement is combind with anomalous dispersion.

In this thesis work, molecular replacement and experimental phasing by SAD and MAD techniques were employed to solve the structures of different proteins. Molecular replacement was carried out with PHASER (McCoy et al., 2007) and experimental phasing by SAD or MAD was performed using the AUTOSHARP pipeline (Vonrhein et al., 2007), which incorporates SHELX (Sheldrick, 2008) for locating heavy atom sites, SOLOMON (Abrahams and Leslie, 1996) for solvent correction and ARP-WARP for automated density

interpretation (Langer et al., 2008). Manual adjustments of crystallographic models were performed with COOT (Emsley et al., 2010) and refinement with REFMAC5 (Murshudov et al., 2011). Validation of models were carried out with MOLPROBITY (Chen et al., 2010), interaction properties, including buried surface area, were calculated with PISA (Krissinel and Henrick, 2007) as implemented by the European node of the Protein Structure Database (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver). Structural homologues were identified using DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server/; Holm and Rosenstrom (2010)) and structure superimpositions were done with CCP4MG (McNicholas et al., 2011) and PyMOL (Schrödinger, LLC). Structure based sequence alignments were performed with with MULTISEQ (Roberts et al., 2006) and phylogenetic trees were generated with the MEGA 6 package (Tamura et al., 2013).

3.3 Snake adenovirus 1 fibre protein

3.3.1 Viral DNA isolation

Snake adenovirus 1 propagated in iguana heart cells (IgH-2; ATCC reference CCL-108; Sorensen and Mesner (2005)) and purified in double caesium chloride gradients as described (Menendez-Conejero, 2013) was obtained from Rosa Menendez-Conejero, Gabriela Condezo-Castro and Carmen San Martín (Dpto. de Estructura de Macromoléculas, CNB-CSIC). The final virus concentration was 3.2×10^{10} viral particles per millilitre of 0.15 M sodium chloride buffered with sodium 4-(2-hydroxyethyl)-1-piperazineethanesulfonate (HEPES) pH 7.8 (as quantified by absorbance measurements; Maizel et al. (1968)). Proteinase K (New England Biolabs, Ipswich, Massachusetts, USA) was added to a final concentration of 1 mg/ml to 0.15 ml virus solution. The tube was incubated overnight at 37 °C. The next day, 0.15 ml water was added and DNA was extracted with 0.3 ml phenol:chloroform:isoamyl alcohol [25:24:1 (v:v:v)] saturated with 10 mM Tris–HCl pH 8.0, 1 mM EDTA (Sigma–Aldrich, St Louis, Missouri, USA) and then with chloroform:isoamyl alcohol [24:1 (v:v)]. DNA was precipitated by addition of sodium chloride to a final concentration of 0.4 M and absolute ethanol [5:2 (v:v) ethanol:sample ratio] and incubation overnight at -20 °C. Following centrifugation (10 min at 16100 x *g*), the DNA pellet was washed with 0.3 ml 70% (v/v) ethanol. Approximately 0.6 μg of viral DNA was obtained.

3.3.2 Construction of expression vectors

Expression constructs for full-length and N-terminally truncated versions of snake adenovirus 1 (SnAdV-1) fibre protein were designed, the putative C-terminal head domain likely being important for trimer assembly and correct folding. Truncations were made taking into account secondary structure prediction (PSIPRED server at http://bioinf.cs.ucl.ac.uk/psipred/; McGuffin et al. (2000)) and potential shaft domain repeats (van Raaij et al., 1999b).

DNA fragments for constructs starting at different residues (1, 82, 171, 234, 293 and 343) were amplified by PCR using the oligonucleotides shown in Table 3. Fragments were digested with restriction enzymes *Bam*HI and *Hind*III and cloned in the pET28c(+) expression vector previously treated with the same enzymes. The expression constructs were named pET-28c-SnAdV1f1, pET-28c-SnAdV1f82, pET-28c-SnAdV1f171, pET-28c-SnAdV1f234 and pET-28c-SnAdV1f293, respectively. DNA Sequencing (Secugen SL, Madrid, Spain) of recombinant plasmids revealed a major discrepancy between the database sequence and the cloned sequences. An insertion of two bases, which results in a change of the five amino acids beyond position 339 and a termination codon (TGA) after position 345 was evident (Singh et al., 2013b). All constructs contained an N-terminal 6xHis tag.

Table 3: SnAdV1 fibre protein fragments and corresponding oligonucleotide used for coding segment amplification.

Fragments	Oligonucleotide forward	Oligonucleotide reverse
1-345	f1-ForBamHI (5′-ggtcggatcctgaagaagataaaacgatcagcggc-3′)	SnAdVf-RevHind (5′-gatcccaagctttcagcttaactcgggacattcttg-3′)
82-345	f82-ForBamHI (5′-ggtcggatccctcctcttcagcaaagcgg-3′)	
171-345	f171-ForBamHI (5′-ggtcggatcctggagtcctatcctcttcccc-3′)	
234-345	f234-ForBamHI (5′-ggtcggatccccagcccgccttctaagac-3′)	
293-345	f293-ForBamHI (5′-ggtcggatcctctcttctttcctcactgccc-3′)	
343-345	f343-ForBamHI (5′-ggtcggatcctgaaagtgatatccatttatttcatggac-3′)	

3.3.3 Site-directed mutagenesis in expression construct pET-28c-SnAdV1f234

In order to allow phasing by MAD on incorporated selenium atoms, two leucine codons at positions 322 and 324 were mutated to methionine codons in plasmid pET-28c-SnAdVf234. A single set of primers (5′-GAC GAG GGA ATT **ATG** ACT **ATG** GAG ATT TCT CGC-3′ and 5′-GCG AGA AAT CTC **CAT** AGT **CAT** AAT TCC CTC GTC-3′) was designed, containing the desired leucine to methionine changes shown in bold. A 50 µl reaction mixture containing 100 ng pET-28c-SnAdV1f234 plasmid and *Pfu* Turbo DNA polymerase (Agilent Technologies, Santa Clara, California, USA) was subjected to 16 cycles of PCR with a final extension step of 7 min. For digestion of parental plasmid DNA, the PCR product was treated with *Dpn*I (Agilent Technologies). Fourty units of *Dpn*I were used, where a unit is defined as the amount required to digest 1 µg of Dam-methylated pBR322 DNA in 1 h at 37 °C. The reaction was incubated for 3 h at 37 °C. The two introduced mutations were confirmed by DNA sequence analysis (Secugen SL, Madrid, Spain) and the resulting vector was named pET-28c-SnAdV1f234mut.

3.3.4 Expression and purification of SnAdV1f1, SnAdV1f82, SnAdV1f171 and SnAdV1f234

E. coli BL21(DE3) cells were transformed with respective plasmids and four 700 ml cultures were grown. After lysis, target proteins were purified from the supernatant by Ni-NTA chromatography and dialysed against buffer containing 10 mM Tris-HCl pH 8.5 for strong anion-exchange chromatography using a Resource Q column. The protein was eluted with a linear gradient of 0–1 M sodium chloride in 10 mM Tris-HCl pH 8.5.

3.3.5 Expression and purification of f234-345mut with selenomethionine

Selenomethionine-derivatised protein (from the construct starting at residue 234 in which leucines 322 and 324 were mutated to methionine) was prepared by transforming methionine auxotroph *E. coli* B834(DE3) cells with pET-28c-SnAdV1f234mut. A 50 ml culture in LB of transformed bacteria were harvested by centrifugation (5000 x *g*, 10 min) and resuspended in 100 ml SelenoMet medium (Molecular Dimensions, Newmarket, Suffolk, England) supplemented with methionine. This culture was grown at 37 °C until the O.D. reached 0.2 as measured at 600 nm and then used as inoculum for 1.6 l SelenoMet medium (Molecular Dimensions, Newmarket, Suffolk, England), supplemented with selenomethionine. After growth and induction at a measured O.D. of between 0.6 and 0.8 at 600 nm, the culture was

allowed to grow overnight at 16 °C. The same cell disruption and protein purification procedures were followed as performed in the case of the native protein construct starting at residue 234, except for the addition of 10 mM beta-mercaptoethanol to all buffers. The obtained protein (234-345Semet) was desalted and concentrated, keeping 10 mM beta-mercaptoethanol in the final buffer (10 mM Tris-HCl pH 8.5, 10 mM beta-mercaptoethanol).

3.3.6 Crystallization of fragments (171-354), (234-354) and (234-345Smet)

Initial crystallization trials for the native protein construct starting at residue 234 were carried out using sitting-drop vapour diffusion. Commercially available kits were employed at 277 and 294 K using either robotic or manual setups. In robotic setups, a Genesis RSP 150 workstation (Tecan, Männedorf, Switzerland) was used, with MRC 96-well sitting-drop crystallization plates (Molecular Dimensions, Newmarket, Suffolk, England). Reservoir volumes were 50 µl and drops were prepared by mixing 0.2 µl protein solution with 0.2 µl reservoir solution. For manual setups, CompactClover plates (Jena Bioscience GmbH, Jena, Germany) were used, with reservoir volumes of 0.15 µl and drops prepared by mixing 1 µl protein solution with 1 µl reservoir solution. The crystals used for data collection were always those from manual setups.

Soaking with several heavy atoms (Hg, Au, Pb, Zn, Br, I) were carried out in original mother liquor except for one instance were native 234 crystals were transferred to 3 M sodium malonate pH 5.0 (Holyoak et al., 2003).

X-ray diffraction data for all the crystals were collected at the European Synchrotron Radiation Facility (ESRF), Grenoble, France or the Diamond Light Source (DLS), Harwell, England at 100 K. Data collected at the DLS were integrated using XIA2 (Winter et al., 2013), which incorporates XDS (Kabsch, 2010) and AIMLESS (Evans and Murshudov, 2013). Structure solution with selenomethionine derivative multi-wavelength anomalous dispersion (MAD) data was performed with the AUTOSHARP pipeline (Vonrhein et al., 2007). Reflections for calculating the free R-factor were selected in thin shells and the same reflections were selected for the free and working R-factor for refining the selenomethionine and native model (both in $I2_13$ space group). A similar approach was adopted to select reflections in thin shells for both $P2_12_12_1$ space group crystals but independently from each other. Reflections were selected randomly in case of $F23$ crystal. Structure solution for the $P2_12_12_1$ data sets was performed by molecular replacement using PHASER (McCoy et al., 2007), taking the previously determined solution as a search model.

3.4 Turkey adenovirus 3 (TAdV-3) fibre protein

3.4.1 Construction of expression vectors

Based on secondary structure prediction of the fibre protein of TAdV-3, two expression constructs were prepared by Mónika Z. Ballmann in the group of Dr. Balázs Harrach and Dr. Mária Benkő at Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary. One of the constructs was for the full-length protein (residues 1-454) while the other was meant to express putative C-terminal fibre head domain (residues 304-454). The respective coding segments were amplified from a commercially available turkey hemorrhagic enteritis vaccine containing attenuated live avirulent virus and cloned into the pET28a(+) expression vector, between the *Eco*RI and *Hin*dIII restriction sites. The constructs were named pET28a(+)-THEVfib1-454 and pET28(+)-THEVfib304-454. The foreign proteins expressed in this vector have an N-terminal tag (M GSSHH HHHHS SGLVP RGSHM ASMTG GQQMG RGSEF) for easy purification by nickel-affinity chromatography.

3.4.2 Site-directed mutagenesis to generate virulent fibre head protein of TAdV-3

Two sets of mutagenic oligonucleotides were designed to generate the mutations in a two-step procedure (Table 4). Met354 to Ile mutation was generated adopting methodology described in section 3.2.2.4. Parental DNA was digested by *Dpn*I treatment and the product was gel purified using GeneJET gel extraction kit (Thermo Fisher, Pittsburgh, USA). TOP10 cells were transformed with the PCR product and grown in 10 ml LB culture. Following plasmid purification the mutation was confirmed by sequencing (Secugen SL, Madrid, Spain) and the plasmid was named pET28a(+)-THEVfib304-454_mutM2I. In order to generate the double mutant (M354I and M376T), the pET28a(+)-THEVfib304-454_mutM2I plasmid was subjected to the same mutagenic PCR procedure but with the oligonucleotide set containing Met376 to Thr mutation. Mutation was confirmed by sequencing and the recombinant plasmid was called pET28a(+)-THEVfib304-454_DoubleMut expressing the THEVfib304-454_virulent fibre head protein.

Table 4: Mutagenic primers to generate virulent version of TAdV-3 fibre head domain.

Mutation	Oligonucleotide forward	Oligonucleotide reverse
Met to Ile	f304-For-Mut_I (5′-gataacataggtgtaatagaaaaccctacc-3′)	f304-Rev-Mut_I (5′-ggtagggttttctattacacctatgttatc-3′)
Met to Thr	f304-For-Mut_T (5′-gctgatttcttgagtcctacgttaaataatacatatatagtgcca-3′)	f304-Rev-Mut_T (5′-tggcactatatatgtattatttaacgtaggactcaagaaatcagc-3′)

3.4.3 Site-directed mutagenesis to generate five single amino acid mutants in the fibre head fragment fib(304-454)

Site-directed mutagenesis was performed to generate five separate single amino acid mutants in the fibre head protein of avirulent TAdV-3. The mutations were: Arg368Ala, Glu392Ala, Asn407Ala, Lys421Ala and Lys439Ala. The pET28a(+)-THEVfib304-454 recombinant plasmid was used as the template, while five sets of oligonucletides were used in separate reactions to introduce the desired mutations (Table 5). Reaction conditions were same as described in section 3.2.2.4 and above.

Table 5: Mutagenic primers to generate single point mutants of avirulent TAdV-3 fibre head domain.

Mutation	Oligonucleotide forward	Oligonucleotide reverse
Arg368Ala	5′-agtattgaattagcatctgctgatttcttg-3′	5′- caagaaatcagcagatgctaattcaatact-3′
Glu392Ala	5′-ggggtaagggtagcatcacctactattcct-3′	5′-aggaatagtaggtgatgctacccttacccc-3′
Asn407Ala	5′-atacttgaaaacgcttcctctttcattcaagtaggg-3′	5′-ccctacttgaatgaaagaggaagcgttttcaagtat-3′
Lys421Ala	5′-aggttaacagttgcgaatggtaaccctcatatg-3′	5′-catatgagggttaccattcgcaactgttaacct-3′
Lys439Ala	5′-cctgggaatattgcaatgataaagataaaatctgtaatg-3′	5′- cattacagattttatctttatcattgcaatattcccagg-3′

3.4.4 Expression and purification of putative fibre head fragment fib(304-454)

For expressing and purifying putative fibre head domain of avirulent TAdV-3, *E. coli* BL21(DE3) was transformed with the recombinant plasmid pET28a(+)-THEVfib304-454. After cell growth and lsyis, the protein was purified from the supernatant by Ni-NTA chromatography. Fractions containing the target protein were pooled and dialysed three times against 500 ml buffer containing 25 mM MES-NaOH pH 6.0 using dialysis tubing of 10 kDa MW cutoff. The dialysed sample was loaded onto a 6 ml Resource S column and eluted with a 0-1 M sodium chloride gradient in the same buffer. The protein eluted at around 0.3-0.4 M sodium chloride. Fractions containing pure protein were pooled and concentrated to 11 mg/ml using Amicon Ultra concentrators, incorporating three washes with 25 mM MES pH 6.0 to remove sodium chloride. Protein was stored temporarily at 4 $^{\circ}$C prior to use in crystallization trials.

The same procedure and maintaining similar buffers as above was adopted for expressing and purifying the putative fibre head domain of virulent TAdV-3 as well as for all five single amino acid mutants of the avirulent fibre head protein.

3.4.5 Expression and purification of avirulent fib(304-454) with selenomethionine

To prepare selenomethionine-derivatised protein, a 50 ml culture of pET28a(+)-THEVfib304-454 transformed B834(DE3) cells was grown overnight at 37 °C in LB media supplemented with kanamycin. Next day, cells were harvested by centrifugation (5000 g, 10 min) and resuspended in 100 ml SelenoMet medium (Molecular Dimensions, Newmarket, Suffolk, England) supplemented with methionine. This culture was allowed to grow at 37 °C until O.D. reached 0.2 at 600 nm and then used as inoculum for 1.6 l SelenoMet medium (Molecular Dimensions, Newmarket, Suffolk, England) supplemented with selenomethionine. Following growth at 37 °C, once an O.D. of 0.6-0.8 had been achieved, the cultures were cooled on ice and induced with IPTG, to be followed by overnight growth at 16 °C. The cell disruption and protein purification procedures were the same as for the native protein except for the addition of 7-10 mM beta-mercaptoethanol to all buffers to prevent oxidation of selenomethionine residues. The selenomethionine-derivatised protein was desalted and concentrated for crystallization in the same way as the native protein, but including beta-mercaptoethanol.

3.4.6 Crystallization, data collection and data processing of avirulent fibre head, avirulent fibre head with selenomethionine and virulent fibre head

To obtain suitable crystallization conditions, commercial screens from several suppliers were tested at 4 and 21 °C either by manual or robotic setups. Reservoir volumes were kept 0.15 µl and drops were made by mixing 1 µl each of protein and reservoir solution. Based on initial hits, a condition containing 1.0 M diammonium phosphate, 0.1 M sodium citrate pH 5.5 was chosen for optimization in sitting-drop vapour-diffusion crystallization experiments in CombiClover 4 Chamber plates (Jena Bioscience, Jena, Germany).

Soaking experiments on crystals of avirulent native fib304-545 protein were performed with 2,3'- and 2,6'-sialyllactose sugars (Elicityl, Crolles, France) at 21 °C. Crystals were soaked between 2 h to overnight in their original growth condition with a sugar concentration of 30-50 mM. Crystal harvesting and mounting were done in same way as for the native crystals.

For data collection, crystals were transferred to a solution containing the same components as the reservoir solution plus 25% (v/v) glycerol and soaked for 30 s. Data was collected on beamline ID14-4 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France using an ADSC Q315r CCD detector and on BL13-XALOC beamline at ALBA synchrotron light source, Barcelona, Spain using a Pilatus 6M (Dectris) detector. Structure solution with Se-derivative data set was performed with the AUTOSHARP pipeline (Vonrhein et al., 2007), employing SAD. As the data indicated presence of only one monomer in the asymmetric unit, reflections for calculating the free R-factor were selected randomly and the same reflections were selected for the free and working R-factor for refining the selenomethionine and native models. For refinement of ligand-bound models, the same set of reflections was used for the free and working R-factor as were selected and used for the native model.

3.4.7 Thermofluor protein unfolding assay

This assay was performed to determine melting temperatures of proteins and was conducted in the presence of fluorescent dye SYPRO Orange (Life Technologies SA, Madrid, Spain) that can bind to hydrophobic patches on protein surface (Lavinder et al., 2009). Upon gradual increment in temperature a protein starts to unfold and more and more hydrophobic patches become available for the dye to bind, giving rise to an increment in fluorescent signal. At one temperature protein exists in equally populated folded and unfolded states, which is reffered to as the melting point (Tm) for that particular protein.

Thermal shift assays on fibre head proteins were carried out in an iCycler iQ PCR Thermal Cycler (Bio-Rad, Hercules, CA, USA). 30 µl reaction volumes were prepared in 200 µl capped PCR tubes by mixing fluorescent dye SYPRO Orange (Life Technologies SA, Madrid, Spain) with 30 µg of protein, maintaining final dye concentration of 1x. Thermal denaturation curves and melting temperatures (Tm) values were obtained by heating the samples through 4 ºC to 94 ºC with a ramp rate of 1 ºC/min. The fluorescence in each well was measured at regular intervals of 0.5 ºC.

3.4.8 Surface Plasmon resonance (SPR)

Surface plasmon resonance experiments were carried out on a BIACORE 3000 (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) platform in order to determine binding kinetics of virulent and avirulent TAdV-3 fibre head proteins with 2,3'-sialyllactose. The Streptavidin sensor chip (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) was prepared

by consecutive injection of solution containing 1 M sodium chloride and 50 mM NaOH prior to ligand immobilization. Biotinylated glycan 2,3'-sialyllactose-sp-biotin was procured (Elicityl, Crolles, France) and passed in varying amount over the chip to attain suitable ligand capture level, which later was set to 150 RU for all assays. Free biotin binding sites were blocked by injecting 100 μM of D-biotin. One of the flow cells of the chip was kept blank and used as reference.

All interaction assays were conducted at 25 ºC. HBS-EP (0.01 M *N*-2-hydroxyethylpiperazine-*N*-2-ethanesulphonic acid-NaOH pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20; GE Healthcare Bio-Sciences AB, Uppsala, Sweden) was used as running buffer during the operation and flow cell surface regeneration was performed with 10 mM glycine-HCl pH 2.0 solution, whenever required. Protein samples were prepared in running buffer and were injected at 30 μl/min in a broad concentration range of 10 μg/ml to 240 μg/ml. The association phase between analyte and ligand was monitored for 8 minutes whereas their dissociation was recorded for 4 minutes. Obtained data were processed with BIACORE 3000 control software (GE Healthcare Bio-Sciences AB, Uppsala, Sweden).

3.5 Lizard adenovirus-2 short (fibre-1) and long (fibre-2) fibres

3.5.1 Construction of expression vectors

Expression constructs for lizard adenovirus-2 short (fibre1; 331 amino acids) and long (fibre2; 433 amino acids) fibres were supplied by Mónika Z. Ballmann, Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary. The inserts were cloned in pET28a(+), between the *Eco*RI and *Hind*III restriction sites, for expression with an N-terminal tag (M GSSHH HHHHS SGLVP RGSHM ASMTG GQQMG RGSEF). While designing the constructs, the C-terminal end carrying the putative fibre head domain was preserved and truncations were done on the N-terminal side. All prepared constructs were checked by restriction analysis for presence of insert as well as by DNA sequencing using T7 forward and reverse primers. Different constructs are given in Table 6.

Table 6: Expression constructs and respective proteins of lizard adenovirus 2 short and long fibres.

Fibre1 constructs	Respective protein	Fibre2 constructs	Respective protein
pET28a(+)-Liz-F1-1	fib1(1-331)	pET28a(+)-Liz-F2-1	fib2(1-433)
pET28a(+)-Liz-F1-46	fib1(46-331)	pET28a(+)-Liz-F2-40	fib2(40-433)
pET28a(+)-Liz-F1-160	fib1(160-331)	pET28a(+)-Liz-F2-118	fib2(118-433)
pET28a(+)-Liz-F1-220	fib1(220-331)	pET28a(+)-Liz-F2-245	fib2(245-433)
pET28a(+)-Liz-F1-241	fib1(241-331)	pET28a(+)-Liz-F2-305	fib2(305-433)

3.5.2 Expression and purification of fib1(220-331)

I decided to express the protein at 16 °C based upon its solubility behaviour. BL21(DE3). Supernatant obtained from four 700 ml LB cultures was subjected to Ni-NTA chromatography in order to separate out the target protein. The fractions containing pure protein, as indicated by SDS-PAGE, were pooled and dialysed three times against 10 mM Tris–HCl pH 8.5 for strong anion-exchange chromatography using a Resource Q column. Elution was performed with a 0–1 M sodium chloride gradient in the same buffer. Fractions containing pure protein were pooled and concentrated using Amicon Ultra concentrators, incorporating three washes with 10 mM Tris–HCl pH 8.5 to remove sodium chloride and stored temporarily at 4 °C prior to use in crystallization trials.

3.5.3 Crystallization, data collection and data processing of fib1(220-331)

For initial trials, four commercially available kits were employed, in total testing 250 conditions at 4 and 21 °C either by manual or robotic setups, keeping the total drop volume at 2 µl or 1 µl. Two conditions produced crystals: one containing 20% (w/v) PEG 3350 and 0.2 M tri-potassium citrate, while the second contained 0.1 M Tris-HCl pH 8.0 and 30% (v/v) PEG 600. The PEG 3500 concentration in the drop was gradually increased to 35% prior to crystal harvesting in the PEG 3500 condition while crystals were mounted directly from the PEG 600 condition. X-ray diffraction experiments were conducted and data sets were obtained at 100 K on beamline ID29 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France and on Proxima-1 beamline of the Soleil synchrotron, Paris, France, in both occasions using a Pilatus 6M (Dectris) detector. Reflections were taken in thin shells for calculating the free R-factor and the same reflections were selected for the free and working R-factor for refinement purpose.

3.5.4 Expression and purification of fib2(245-433)

Four 700 ml cultures in LB media were inoculated with a starter culture of pET28a(+)-Liz-F2-245 transformed C41(DE3) cells in a v/v ratio of 1:1000. Cells were lysed, supernatant was collected and subjected to Ni-NTA chromatography for purification of the target protein. Collected fractions were checked for the presence of target protein, pooled together and dialysed three times against anion exchange chromatography buffer (10 mM Tris–HCl pH 8.5) using dialysis tubing of 10 kDa cutoff. A Resource Q column was used for anion exchange chromatography and elution was performed with a 0-1 M sodium chloride gradient in the same buffer. Fractions were pooled and concentrated prior to use in crystallization trials.

3.5.5 Crystallization, data collection and data processing of fib2(245-433)

Four commercially available kits were tested and only one condition consisting of 0.2 M di-ammonium tartrate and 20% w/v PEG 3500 could yield crystals. Crystals were harvested after gradually increasing the PEG 3500 concentration to 35% in order to prevent freezing. X-ray diffraction data was collected on Proxima-1 beamline of the Soleil synchrotron, Paris, France using a Pilatus 6M (Dectris) detector.

Automatic data processing was performed with XIA2 pipeline, which incorporates XDS (Kabsch, 2010) and AIMLESS (Evans and Murshudov, 2013).

3.6 Murine adenovirus-2 (MAdV-2) fibre protein

3.6.1 Construction of expression vector

Three expression constructs were made in pET28a(+) vector by Márton Vidovszky of the group of Dr. Balázs Harrach at the Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary. Constructs design was based on secondary structure prediction (PSIPRED server at http://bioinf.cs.ucl.ac.uk/psipred/); one full length (1-787 residues) and two N-terminal truncated versions (517-787 and 586-787 residues) were prepared. All inserts were cloned between *Eco*RI and *Hind*III cloning sites. Residues 586-787 were presumed to compose the fibre head domain. Details of construct and respective encoded proteins are provided in Table 7.

Table 7: Expression constructs and respective proteins of MAdV-2 fibre.

Construct	Encoded protein
pET28a(+)-MAdV2_fib1	fib(1-787)
pET28a(+)-MAdV2_fib517	fib(517-787)
pET28a(+)-MAdV2_fib586	fib(586-787)

3.6.2 Expression and purification of fib(517-787) and fib(586-787)

The MAdV-2 fib(517-787) fibre protein fragment was expressed in BL21(DE3) cells. Ni-NTA purification was performed to separate out the target protein from the supernatant. Presence of target protein in different fractions was determined by SDS-PAGE analysis. Fractions containing protein were pooled together to be dialysed against buffer containing 10 mM Tris-HCl pH 8.5 using dialysis tubing of 10 kDa MW cutoff prior to anion exchange chromatography. Post anion exchange chromatography purification, during which the protein was eluted using a linear gradient of 0–1 M sodium chloride, fractions containing pure protein were pooled together, washed and concentrated using Amicon Ultra concentrators in buffer consisting of 10 mM Tris-HCl pH 8.5. The MAdV-2 fibre protein fragment fib(586-787) was also purified in a similar manner using the same buffers. Both proteins were stored temporarily at 4 °C prior to further utilization.

3.6.3 Crystallization, data collection and data processing of MAdV-2 fibre protein constructs

For crystallizing the fib(586-787) protein, around 350 conditions were tested at 21 °C, in manual as well as in robotic setup, to obtain initial hits. Drop volume was kept 1 µl (0.5 µl protein + 0.5 µl reservoir solution) or 2 µl (1 µl protein + 1 µl reservoir solution). Optimization of initial hits resulted in plate like crystals, which appeared in a condition containing 5M ammonium formate and HEPES-NaOH pH 7.5. Crystals from this condition were mounted and vitrified directly. Two more conditions consisting of 0.1 M magnesium chloride, 0.1 M HEPES-NaOH pH 7.5 and 10 % w/v PEG 4000 and 0.15 M ammonium sulphate, 0.1 M HEPES-NaOH pH 7.0, 20 % w/v PEG 4000 also yielded crystals. These crystals were first transferred to a solution containing the same components as the reservoir solution plus 20% (v/v) glycerol and soaked for 30 s prior to mounting and vitrification.

For heavy atom derivatisation, methyl mercury chloride (CH_3HgCl) solution was added to the reservoir in a final concentration of 5 mM and left overnight to equilibrate the osmotic

balance of the chamber. 2 µl reservoir solution was then added to the drop and crystals were soaked between 30 s to 1 min.

X-ray diffraction data sets were collected at 100 K on BL13-XALOC beamline at ALBA synchrotron, Barcelona, Spain using a Pilatus 6M (Dectris) detector. The AUTOSHARP version 3.10.2 pipeline (Vonrhein et al., 2007) was used to determine the structure from a single mercury derivative data set employing single-wavelength anomalous dispersion. Due to presence of non-crystallographic symmetry, reflections for calculating the free R-factor were selected in thin shells and the same reflections were selected for the free and working R-factor for refining the derivative and native models.

3.6.4 Expression and purification of the human coxsackievirus and adenovirus receptor (hCAR)

Expression of the D1 domain of the human coxsackievirus and adenovirus receptor (hCAR) was mostly based on a method described previuosly previously (van Raaij et al., 2000). A pAB3 expression vector containing the hCAR-D1 domain (residues 18-140) coding segment and a C-terminal 6xHis tag was used to tranform XL-10 gold cells. The cells were then allowed to grow aerobically at 25 °C in two 700 ml cultures until the O.D. of 0.6-0.8 measured at 600 nm was attained, thereafter induced with 0.2 mM IPTG and left for further 16 hours with shaking incubation at the same temperature. Cells were harvested by centrifugation at 5000 x *g* for 15 min and resuspended in lysis buffer consisting of 20 mM Tris–HCl pH 8.0, 300 mM sodium chloride, 10 mM imidazole, 5% (v/v) glycerol and a Complete protease inhibitor cocktail tablet, EDTA free (Roche, Indianapolis, USA). Lysis was carried out in a French press (three passes) and the cell debris was removed by centrifugation at 20000 x *g* for 45 min resulting in a clear supernatant. The supernatant was incubated with 1 ml Ni–NTA resin (Jena Bioscience GmbH, Jena, Germany) at room temperature for 1 h and then loaded onto an empty column, followed by washing twice with 10 column volumes of buffer containing 20 mM Tris–HCl pH 8.0, 300 mM sodium chloride, 20 mM imidazole, 5% (v/v) glycerol. Step-gradient elution of the bound protein was performed using imidazole (50, 100, 200, 300, 400 and 1000 mM) in the same buffer. Elution fractions were checked by SDS-PAGE and pooled together. Dialysis against buffer consisting of 10 mM Tris-HCl pH 8.5 was performed using 10 kDa MW cutoff dialysis tubing to remove imidazole and sodium chloride followed by anion exchange chromatography using a Resource Q column in which the protein was eluted with a linear gradient of 0-1 M sodium

chloride. Pure protein-containing fractions were pooled together, washed and concentrated using Amicon Ultra concentrators in 10 mM Tris-HCl pH 8.0.

3.6.5 Affinity assay with MAdV-2 fibre head

A gel filtration assay was performed to analyze binding of hCAR to the MAdV-2 fibre head. Both proteins (in 10 mM Tris-HCl pH 8.0, 150 mM sodium chloride) were mixed in a trimer to monomer molar ratio of 1:3 and incubated for 30 mins at 4 °C. A 25 ml Superdex 200 GL gel filtration column (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) was prepared with buffer containing 10 mM Tris-HCl pH 8.0 and 150 mM NaCl and the mix was loaded onto it. Elution was performed with 1 column volume of the same buffer as used for column equilibration and peaks were analysed by SDS-PAGE to detect presence of proteins.

3.7 Snake adenovirus-1 (SnAdV-1) LH3 protein

3.7.1 Construction of the expression vector

Coding sequence for 373 amino acid residues long LH3 protein was amplified with a set of primers (ForBamHI-LH3: 5′-gggggatccatgacttctgttgaagat-3′ and RevNotI-LH3: 5′-attatagcggccgctcatgcggcgggctg-3′) adding a forward *Bam*HI and reverse *Not*I restriction sites to the insert by a 3 step PCR reaction using genomic DNA of SnAdV-1 as a template (kindly provided by Dr. Carmen San Martin, CNB-CSIC, Madrid, Spain). The insert was cloned in pET28a(+) previously digested with same set of restriction enzymes. The protein was expressed with an N-terminal 6XHis tag. Insert sequence was confirmed by DNA sequencing (Secugen SL, Madrid, Spain) and recombinant vector was named pET28a(+)-LH3.

3.7.2 Protein expression and purification

Four 700 ml cultures in LB medium were inoculated with a starter culture of pET28a(+)-LH3 transformed JM109(DE3) cells. Cells were grown, harvested and lysed as previously described. Target protein was purified from the collected supernatant by Ni-NTA chromarography and fractions containing the target protein were collected and dialysed against buffer consisting of 10 mM Tris-HCl pH 8.5 using 10 kDa dialysis tubing. This is followed by anion exchange chromatography using a Resource Q column. After this step, protein was washed and concentrated using Amicon Ultra concentrators in 10 mM Tris-HCl pH 8.5 and stored temporarily at 4 °C prior to crystallization trials.

3.7.3 Limited proteolysis with alpha-chymotrypsin

Limited proteolysis experiments on LH3 protein were carried out in order to identify a stable domain and also to facilitate crystallization of the protein. Protein samples (2 mg/ml) were incubated at 37 °C, 55 °C and 65 °C for 20 min followed by cooling down to 37 °C. Alpha-chymotrypsin solution (Sigma-Aldrich, St Louis, Missouri, USA) was prepared in 1 mM HCl and 2 mM $CaCl_2$ at 1 mg/ml concentration and added to the samples in a protease to protein ratio of 1:2 and incubated at 37 °C for 20 min. The reaction was stopped by addition of protein gel loading buffer consisting of SDS and Complete protease inhibitor cocktail tablet (Roche, Indianapolis, USA). The results were analyzed by SDS-PAGE gel after which the sample was loaded onto a Resource Q column for separation of proteolysed products.

For *in-situ* proteolysis the previously described method of Dong et al. was adopted (Dong et al., 2007). Alpha-chymotrypsin solution was added to the purified protein on ice immediately prior to crystallization trials in a protease to protein ratio of 1:100. Crystallization was performed in sitting drops at 21°C, adding 0.5 µl of the protease/protein mixture to 0.5 µl of the precipitant solution.

3.7.4 Western blot analysis

The proteolysed protein sample was separated through anion exchange chromatography and 5 µg of the prominent peak fraction was transferred to a nitro-cellulose membrane using wet blotting system (Bio-Rad, Hercules, CA, USA). Membrane blocking was done by incubation with PBS solution containing 5% (w/v) non-fat milk and 0.05% TWEEN 20 for 1 h at room temperature followed by three wash with PBS containing 0.05% TWEEN 20. The protein was detected using Anti-polyHistidine-Peroxidase conjugate antibody (1:2000; Sigma-Aldrich, St Louis, Missouri, USA) in a solution containing PBS, 5% (w/v) non-fat milk and 0.05% TWEEN 20 for 1 h at room temperature. Followed by washing, the membrane was developed using Amersham ECL prime western blotting detection reagent (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) following the instructions of the supplier.

3.7.5 Electron microscopy

Negative staining EM was performed using a JEOL JEM 1230 microscope at 100 KV. The sample was prepared in several dilutions and adsorbed on glow discharged carbon coated grids for 5 minutes. Excess fluid was removed by touching on the edge of the grid with a

piece of Whatman paper, without allowing it to become completely dry. The grids were then incubated the staining agent, 2% uranyl acetate, for 45 sec. After staining, the grids were blotted and air-dried on a filter paper in a Petri dish.

3.7.6 Surface methylation

Surface methylation was tried as a rescue approach to crystallize LH3 protein. The protocol described by Walter et al., which targets surface exposed lysine residues, was adopted (Walter et al., 2006).

3.7.7 Crystallization, data collection and data processing of LH3 protein

Crystallization trials were performed by sitting drop vapor diffusion technique at 4 and 21 $^{\circ}$C, manually as well as with robotic setup. Obtained crystals were also soaked in original reservoir solution supplemented with methyl mercury chloride (CH_3HgCl) in order to obtain heavy atom derivatised crystals for structure solution.

For data collection, crystals were transferred to wells containing same mother liquid plus 20% glycerol. Data was collected on BM30A beamline at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) using an ADSC Q315r CCD detector and on BL13-XALOC beamline at ALBA synchrotron light source, Barcelona, Spain using a Pilatus 6M (Dectris) detector.

Results and Discussion

4. Results and Discussion

4.1 Structure of the C-terminal fibre head domain of snake adenovirus 1

Atadenoviruses have a characteristic gene organization and capsid morphology (Pantelic et al., 2008). Snake adenovirus type 1 (SnAdV-1) contains a single fibre gene that codes for a 345 residue protein. The high-resolution crystal structure of the 107 residues long SnAdV-1 fibre head domain was determined by the multi-wavelength anomalous dispersion (MAD) method. Crystals of four different types were obtained, of which the best diffracted to a maximum resolution of 1.4 Å. This is the first Atadenovirus for which the structure of the fibre head has been determined. Despite the absence of significant sequence homology, this snake atadenovirus fibre head has the same beta-sandwich propeller topology as other adenovirus fibre heads. However, it is about half the size, mainly due to much shorter loops connecting the beta-strands. Glycan microarray profiling showed that the fibre head might bind sialyllactose, which was further confirmed by STD-NMR and ITC, although the binding affinity was low.

4.1.1 Purification and crystallization

4.1.1.1 Construction of expression vectors

Based on primary sequence and structure prediction analysis of the SnAdV-1 fibre protein, six constructs starting at residues 1, 82, 171, 234, 293 and 343 were designed and cloned into the pET28c(+) expression vector. The constructs were designed to express the target protein with an amino-terminal six-histidine purification tag. DNA sequencing of the generated constructs revealed a frame-shift mutation in the fibre gene that created a stop codon just after residue 345. This finding was consistent with the mass spectroscopy data performed on whole virus (Singh et al., 2013b) and hence the products of the last two expression constructs were not detectable by denaturating gel electrophoresis due to their size (Figure 9). Using a one-step site-directed mutagenesis protocol, a double mutant of putative fibre head fragment 234 was generated, in which two leucine residues (323 and 325) were replaced by methionine to facilitate structure solution (Singh et al., 2013b).

```
  1 MKKIKRSAAD PDPVYPFGDE VPIPLPPFLV PGGGLTTDGL SLAVQTVDPL
 51 NVTLGGVGLK IGDGLSVVDG KLTSEAKIVA DPPLQQSGDT LSLSTDSSMM
101 VLPSGQLTIN NLPSISVTSS GVGLVSPNAP LQLMSNGALQ LSVGGGLTVG
151 AQGSLQISTG VGVNVNAAGV LESYPLPPLV WDYSSKSLTL DIGPGLTVVN
201 GKLQVIGATF SNQMSRMAPA PRADLQSNSI EPLPSPPSKT SLDIAEELQN
251 DKGVSFAFQA REEELGAFTK RTLFAYSGDG LTGPFKAPAS AELSSFLTAH
301 PKGRWLIAFP LGTGIVSVDE GILTLEISRS LPEVGSGSSS TSLKVISIYF
351 MDLFFPVPFI DRASHPAPRR SNNSRQLFHS KQRLFLKVKD FKKRSWYSSL
401 FTLINLNIQE CPELS

  1 MKKIKRSAAD PDPVYPFGDE VPIPLPPFLV PGGGLTTDGL SLAVQTVDPL
 51 NVTLGGVGLK IGDGLSVVDG KLTSEAKIVA DPPLQQSGDT LSLSTDSSMM
101 VLPSGQLTIN NLPSISVTSS GVGLVSPNAP LQLMSNGALQ LSVGGGLTVG
151 AQGSLQISTG VGVNVNAAGV LESYPLPPLV WDYSSKSLTL DIGPGLTVVN
201 GKLQVIGATF SNQMSRMAPA PRADLQSNSI EPLPSPPSKT SLDIAEELQN
251 DKGVSFAFQA REEELGAFTK RTLFAYSGDG LTGPFKAPAS AELSSFLTAH
301 PKGRWLIAFP LGTGIVSVDE GILTLEISRS LPEVGSGSSF YLTEK
```

Figure 9: Mass spectrometric analysis of SnAdV1 fibre protein in purified virions. Peptides identified by Mascot in the database sequence (coverage 28%; top) and in the newly determined sequence (coverage 38%; bottom) are shaded grey; the C-terminal peptide identified that is specific to the new sequence is also underlined.

4.1.1.2 Protein purification and crystallization

A two-step purification strategy was adopted to purify the expressed target proteins as described in Materials and Methods. Fragments starting at residues 1 and 82 tend to aggregate in solution, probably due to the presence of a flexible shaft region, whereas protein expression was undetectable in case of fragments 293 and 343. Sufficient amounts of well-behaving protein could only be obtained for constructs containing residues 171-345 and 234-345, from which fragment 234 was presumed to be the fibre head domain. In strong anion-exchange chromatography, the proteins eluted between 0.2 and 0.4 M sodium chloride in two separate peaks (Figure 10). Usually, the first peak was more prominent and was used for crystallization, but in one case crystals were also obtained from the second peak (native form 1, isomorphous crystals obtained from both peaks, pooled independently). This was observed for C-terminal domains for other adenovirus fibres (van Raaij et al., 1999a), but it is not clear to us how the protein differs between the peaks, although charge must somehow be involved. The electrophoretic behaviour under denaturing conditions (observed size of around 16 kDa; Figure 11A) is consistent with the tagged protein construct starting at amino acid 234 and ending at residue 345 (expected monomer size of 15 kDa), but not including 70 extra residues as would have been expected from the database sequence (expected monomer size of 24 kDa). The protein construct starting at amino acid 171 has an observed monomer size of

around 26 kDa (Figure 11B), where 22 kDa would have been expected if it ends at residue 345 or 31 kDa if it ends at residue 415. While for the 234-345 construct pure protein could be obtained, the 171-345 construct contained some impurities as evident from the SDS-PAGE gel (Figure 11B). For both constructs, the protein remains trimeric unless heated to at least 95 ºC for a few minutes before SDS–PAGE. Note the trimer band of the 171-345 construct entered only a few millimeters into the gel, possibly because of the protein being in a slowly migrating 'hydra' form, owing to unfolding of the shaft part in SDS, as previously observed for the human adenovirus type 2 fibre protein (Mitraki et al., 1999).

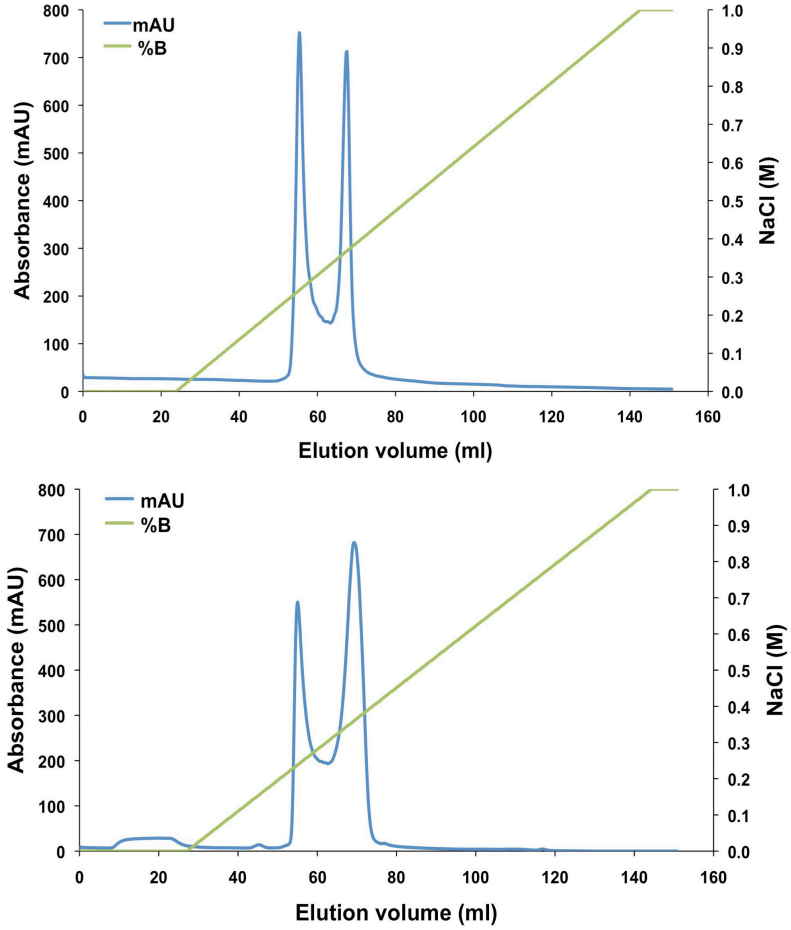


Figure 10: Anion exchange chromatography profiles for the SnAdV-1 construct 171 (top) and SnAdV-1 construct 234 (bottom). In green the sodium chloride concentration gradient is indicated.

Yields of the 234-345 construct were typically around 5 mg of purified and concentrated protein per litre of bacterial culture, while for the 171-345 construct around 6 mg per litre of bacterial culture could be obtained, of which around half was lost upon concentration, indicating that the protein has a tendency to aggregate.

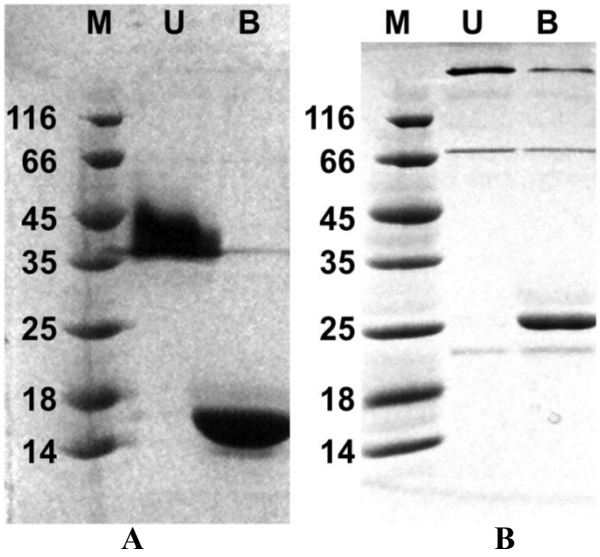


Figure 11: Denaturing gel electrophoresis of purified snake adenovirus 1 fibre protein constructs. Lanes are marked M (molecular weight markers, with the size indicated in kDa), U (unheated sample) and B (sample heated to 368-373 K for at least 5 min). (A) Purified 234-345 construct protein. (B) Purified 171-345 construct protein.

Initial crystallization trials for the native 171-345 and 234-345 protein fragments were performed using sitting-drop vapour diffusion as described in Materials and Methods. Several hundred conditions were tried, of which five conditions produced crystals (Table 8). A solution consisting of 1.7 M ammonium sulfate, 0.085 M HEPES sodium salt pH 7.5, 1.7% (v/v) polyethylene glycol (PEG) 400, 15% (v/v) glycerol, was further selected for optimization in sitting-drop vapour-diffusion crystallization experiments. Crystals of fragment 171–345 of the fibre protein were obtained after about 1 year in 1.5 M ammonium sulfate, 0.075 M Tris–HCl pH 8.5, 25% (v/v) glycerol at 21 $^{\circ}$C. Selenomethionine-derivatised crystals of fragment 234 were also obtained in the same optimization screen around condition consisting of 1.7 M ammonium sulfate, 0.085 M HEPES sodium salt pH 7.5, 1.7% (v/v) polyethylene glycol (PEG) 400, 15% (v/v) glycerol after about 30 days. Obtained crystals are shown in Figure 12 while statistics of the data collected from native and derivative crystals are given in Table 9 and 10 respectively.

Table 8: Crystal producing conditions for SnAdV-1 fibre constructs 171-345 and 234-345

Condition	Composition
1	0.2 M sodium citrate, 0.1 M HEPES-NaOH (pH 7.5), 30% (v/v) 2-Methyl-2, 4-pentanediol
2	0.17 M lithium-sulphate, 0.085 M Tris-HCl (pH 8.5), 25.5% (w/v) PEG 4000, 15% (v/v) glycerol
3	1.26 M sodium citrate, 0.09 M HEPES-NaOH (pH 7.5), 10% (v/v) glycerol
4	1.7 M ammonium sulphate, 0.085 M HEPES-NaOH (pH 7.5), 1.7% (v/v) PEG 400, 15% glycerol
5	1.6 M ammonium sulphate, 0.08 M sodium acetate (pH 4.6), 20% (v/v) glycerol

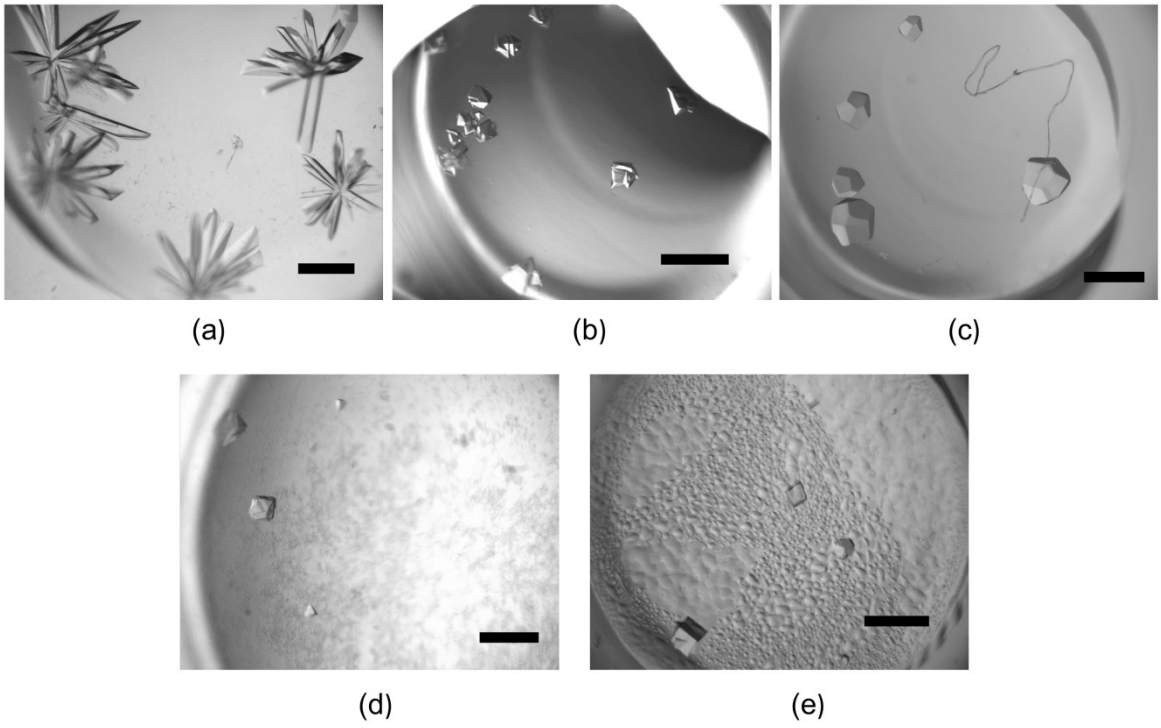


Figure 12: Crystals of snake adenovirus 1 fibre protein constructs. (a) Native crystal form1 (construct 234-345). (b) Native crystal form2 (construct 234-345). (c) Native crystal form3 (construct 234-345). One of the crystals has grown from or around a fibre unintentionally present in the drop. (d) Native crystal form 4 (construct 171-345). (e) Selenomethionine-derivatised crystal of form 3 (construct 234-345). The bars represent 0.5 mm.

Table 9: Crystallographic data measured from native crystals

Data collection	Native 1	Native 2	Native 3	Native 4
Construct (residues)	234-345	234-345	234-345	171-345
Synchrotron, beamline	ESRF, ID14-4[a]	ESRF, ID29[b]	ESRF, ID14-1	DLS, I02
Detector	ADSC Q315r	Pilatus 6M	ADSC Q210	Pilatus 6M
Distance (mm)	386	344	141	271
Wavelength (Å)	0.9393	0.9999	0.9334	0.9797
Number of images	117, 282[c]	3600	100	400
Oscillation range (º)	1.5, 0.4	0.1	0.5	0.15
Space group	$P2_12_12_1$	$P2_12_12_1$	$I23$ or $I2_13$	$F23$
Cell edges (Å, a, b, c)	79.6, 122.3, 133.7	96.8, 96.8, 153.3	149.6, 149.6, 149.6	121.5, 121.5, 121.5
Resolution range (Å)	30.0-2.70 (2.85-2.70)	70.0-1.90 (2.00-1.90)	20.0-1.70 (1.79-1.70)	43.0-1.35 (1.38-1.35)
Reflections	36484 (5266)	113650 (16374)	60984 (8834)	33985 (2424)
Multiplicity	9.3 (9.0)	11.9 (12.1)	6.2 (6.2)	5.8 (2.4)
Completeness (%)	99.7 (99.8)	99.9 (99.9)	99.9 (100.0)	99.5 (95.3)
Mean <I/s(I)>	10.1 (2.8)	19.6 (6.5)	17.3 (4.2)	18.5 (2.2)
R_{sym} (%)[d]	16.7 (87.9)	7.2 (39.0)	6.5 (40.0)	4.8 (37.3)
CC Imean	0.993 (0.794)	0.999 (0.953)	0.999 (0.896)	0.998 (0.732)
CC anom	-0.206 (0.021)	-0.254 (-0.046)	-0.067 (0.010)	0.053 (0.046)
Wilson B ($Å^2$)	55.1	25.1	16.5	15.6

Values in parentheses are for the highest resolution bin.
[a]McCarthy et al., 2009.
[b]de Sanctis et al., 2012.
[c]Data was collected in two different runs
[d]$R_{symm} = \Sigma_h\Sigma_i|I_{hi}-\langle I_h\rangle|/\Sigma_h\Sigma_i|I_{hi}|$, where I_{hi} is the intensity of the *i*th measurement of the same reflection and $\langle I_h\rangle$ is the mean observed intensity for that reflection.

4.1.2 Structure determination of the SnAdV-1 fibre head domain

The structure of the SnAdV-1 fibre head domain was determined up to 1.6 Å resolution by multi-wavelength anomalous dispersion, using data collected from a selenomethionine derivative crystal. Four monomers were present in the asymmetric unit, three (chains A, B and C) forming the biological trimer and the other (chain D) forming a trimer with two crystallographically equivalent monomers in the crystal (space group $I2_13$ has three-fold crystallographic symmetry axes). A total of eight selenium sites (2 per monomer) were expected and all of them were successfully found by SHELX (Sheldrick, 2008). After phase-refinement with SHARP, the overall phasing statistics were excellent, with the most significant contributor to phase information being the dataset collected at the high-energy remote wavelength (overall phasing power of 2.665). Phasing statistics were very good (Table 11). Amino acids from the N-terminal T7 and 6xHis tags could not be modelled in the structure due to lack of density, as was also the case for residues 234-236 of each of the chains and residue Lys345 of chains A, B and C. Using NCS-restraints, the structures of both the seleno-methionine variant and native protein could be refined to low R-values at 1.6 and 1.7 Å resolution, respectively, with good geometry and Ramachandran statistics (Table 11). The head domain, with its eight-stranded beta-sandwich, starts at residue 238.

The structures of $P2_12_12_1$ crystal forms were solved by molecular replacement using a trimer of the determined structure as a template and refined using NCS-restraints at 2.7 and 1.9 Å resolution, respectively. The two different crystal forms belonging to the orthorhombic space group $P2_12_12_1$ both had four complete trimers in the asymmetric unit. The purified 171-345 construct protein crystallized in space group $F23$ after a period of about one year. After structure solution by molecular replacement and refinement of coordinates and individual atom anisotropic temperature factors at 1.33 Å resolution, it was noted that density for residues 171-231 was absent, indicating a possible proteolysis event during crystallization. The asymmetric unit of the crystals contains one monomer of the SnAdV1 fibre head, with all the residues of the fibre head (Pro232 to Lys345) being well-resolved. Refinement statistics for all models are given in table 11.

When the structures of the fibre head monomer of the different crystal forms are superimposed, no significant differences in main chain conformations were observed. The r.m.s.d. between the structures is less than 0.38 Å and even the loop regions superimpose very well, suggesting little flexibility. When trimers are superimposed, the r.m.s.d. is less than 0.5 Å, suggesting the orientation of the monomers in the trimer is also fixed. Given the

structural similarity between the crystal forms, the description below is based on the highest-resolution structure obtained, but is general to all.

Table 10: Crystallographic data measured from a selenomethionine derivative crystal

Data collection			
Construct (residues)	234-345		
Synchrotron, beamline	ESRF, ID14-4[a]		
Detector	ADSC Q315r		
Space group	*I*23 or $I2_13$		
	Inflection point	Peak	Remote
Wavelength (Å)	0.9793	0.9791	0.9768
Distance (mm)	279	279	211
Number of images	200	200	200
Oscillation range (°)	0.5	0.5	0.5
Cell edge (Å, a=b=c)[b]	149.50	149.45	149.54
Resolution range (Å)	30.0-1.90 (2.00-1.90)	30.0-1.90 (2.00-1.90)	30.0-1.60 (1.69-1.60)
Reflections	43740 (6346)	43706 (6352)	72960 (10590)
Multiplicity	12.2 (11.7)	12.2 (11.7)	12.2 (12.2)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
Mean <I/s(I)>	17.5 (8.9)	16.2 (8.5)	18.2 (6.2)
R_{sym} (%)[d]	10.4 (25.4)	11.3 (25.9)	8.8 (40.2)
CC Imean	0.998 (0.981)	0.998 (0.981)	0.999 (0.957)
CC anom	0.500 (0.227)	0.646 (0.403)	0.498 (0.127)
Wilson B ($Å^2$)	14.9	14.7	13.9

Values in parentheses are for the highest resolution bin.
[a]McCarthy et al., 2009.
[b]Although the three datasets were collected from the same crystal, cell parameters were refined independently for all three datasets.
[d]$R_{symm} = \Sigma_h\Sigma_i|I_{hi}-\langle I_h\rangle|/\Sigma_h\Sigma_i|I_{hi}|$, where I_{hi} is the intensity of the *i*th measurement of the same reflection and $\langle I_h\rangle$ is the mean observed intensity for that reflection.

Table 11: Crystallographic structure solution and refinement (1 Å = 0.1 nm)

Phasing					
	Remote	Peak	Inflection point	Overall	
Wavelength	0.9768	0.9791	0.9793		
Resolution range used (Å)	29.3-1.6	29.3-1.9	29.3-1.9		
Reflections (acentric/centric)	69500 / 2391	41317 / 2402	41283 / 2374		
Heavy atom sites				8 Se	
Correlation coefficient (all/weak)				60.91 / 52.64	
Patterson figure of merit				16.56	
Correlation coefficient (E)				0.560	
R-cullis isomorphous (acentric/centric)	0.646 / 0.616	0.686 / 0.665	- / -		
R-cullis anomalous (acentric)	0.641	0.455	0.518		
Phasing power isomorphous (acentric/centric)	1.378 / 0.886	0.864 / 0.466	- / -		
Phasing power anomalous (acentric)	1.899	2.665	2.356		
FOM* (acentric/centric)				0.554 / 0.357	
Solvent flattening#					
R-factor (before/after)				0.3720 / 0.1560	
Correlation on $\|E\|^2$ (before/after)				0.5492 / 0.9063	
Correlation on $\|E\|^2$ / contrast (original / inverted)				0.1248 / 1.2247	
Refinement					
	Selenomethionine	Native 1	Native 2	Native 3	Native 4
Space group	$I2_13$	$P2_12_12_1$	$P2_12_12_1$	$I2_13$	$F23$
Cell parameters (a, b, c, Å)	149.5, 149.5, 149.5	79.6, 122.3, 133.7	96.8, 96.8, 153.3	149.6, 149.6, 149.6	121.5, 121.5, 121.5
Monomers / asymmetric unit	4	12	12	4	1
Resolution (Å)	29.3-1.6 (1.70-1.60)†	29.9-2.7 (2.86-2.70)	29.7-1.9 (2.03-1.90)	29.3-1.7 (1.80-1.70)	43.0-1.33 (1.37-1.33)
Reflections used	70937 (11269)	32882 (5194)	111478 (19922)	59326 (9523)	32152 (2285)
Reflections used for R_{free}	1975 (362)	2057 (329)	2003 (278)	1604 (233)	1832 (138)
R-factor&	0.158 (0.190)	0.212 (0.316)	0.173 (0.212)	0.167 (0.226)	0.117 (0.210)
R-free	0.189 (0.192)	0.247 (0.340)	0.239 (0.250)	0.201 (0.239)	0.138 (0.225)
Number of atoms	3927	9799	10974	4006	1019
Average B / Wilson B ($Å^2$)	18.4 / 13.9	49.5 / 55.1	32.1 / 25.1	19.1 / 16.5	22.1 / 15.6
Ramachandran statistics (%)	98.9/100.0	96.7/99.9	98.5/100.0	98.8/100.0	100.0/100.0
r.m.s.d. bonds (Å)/angles (°)	0.012/1.5	0.016/1.3	0.014/1.4	0.013/1.5	0.013 / 1.6
PDB code	4D0U	4D1F	4D1G	4D0V	4UMI

*Figure of merit = cos(phase error)

#41.5% solvent content

† Values in parentheses are for the highest resolution bin (where applicable).

& $R_{symm} = \Sigma_h\Sigma_i|I_{hi}-\langle I_h\rangle|/\Sigma_h\Sigma_i|I_{hi}|$, where I_{hi} is the intensity of the *i*th measurement of the same reflection and $\langle I_h\rangle$ is the mean observed intensity for that reflection.

4.1.3 Structure of the SnAdV-1 fibre head

The structure of 107-residue SnAdV-1 fibre head revealed a compact head domain (Figure 13C). Despite the low sequence identity with other fibre heads, the topology of the beta-sandwich is identical to that of known adenovirus fibre head structures and contains ABCJ and GHID beta-sheets (Figure 13B; the E- and F-strands in the long DG-loop of human adenoviruses 5 and 2 are absent (Xia et al., 1994; van Raaij et al., 1999a). However, it is significantly smaller. In the trimeric assembly, the ABCH-sheet is partially buried, facing the inner side, while the GHID-sheet is solvent exposed. Given their locations, the former contributes the majority of inter-monomeric contacts whereas latter facilitates intra-monomeric interactions. The strands of the buried sheet are longer in comparison to that of the exposed sheet, as reflected by their average length of 10 residues per strand, *vs* 6 for the strands of the exposed sheet.

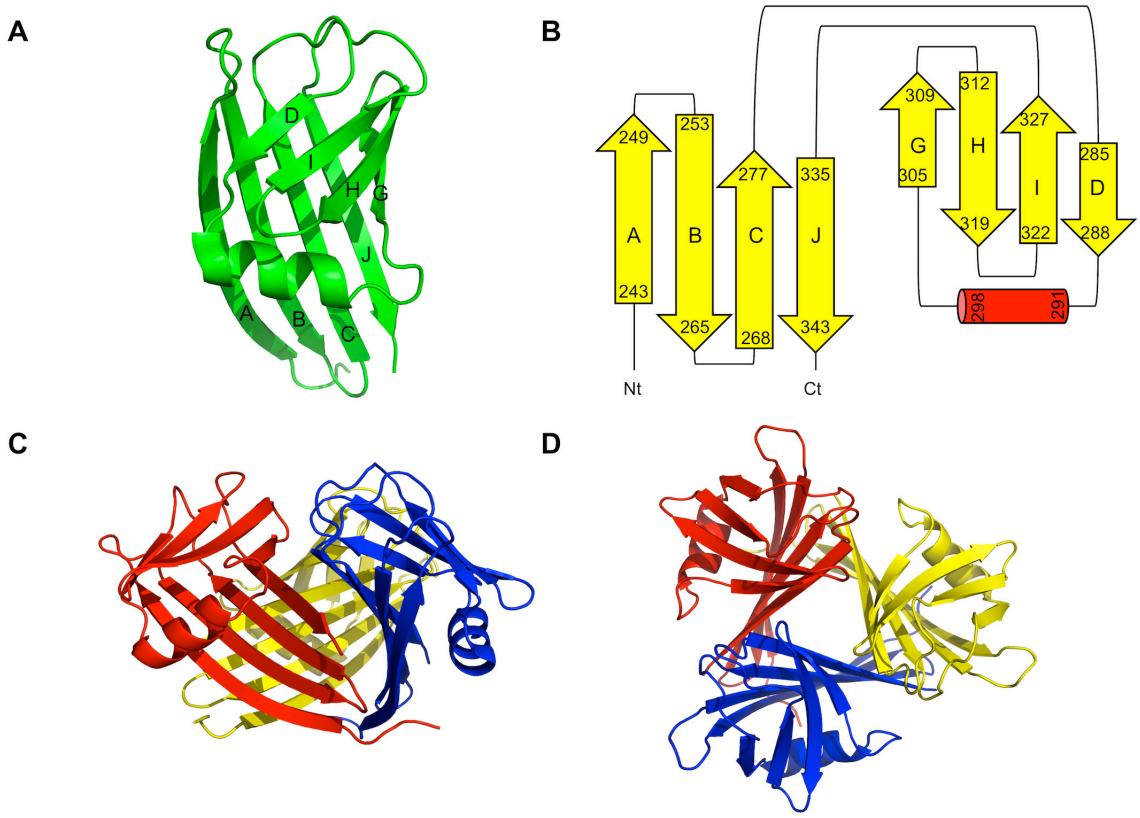


Figure 13: Structure of the snake adenovirus fibre head. (A) Monomer structure. The beta-strands are labeled. (B) Topology. The ABCJ and GHID beta-sheets are coloured yellow, the alpha-helix in the DG-loop is shown in red. The termini are labelled. (C) Trimer structure, side view, with the three monomers colored differently. (D) Trimer structure, top view.

The beta-strands in the structure are connected by loops or by beta-turns. Seven such connections exist; three loops of variable size (CD, DG and IJ) and four beta-turns of 2 residues each (AB, BC, GH and HI). Due to prominence of these tight beta-turns, the overall trimeric assembly is very compact, with a diameter of 4.6 nm and a height of 3.8 nm (compared to a diameter of 6.2 nm and a height of 4.0 nm for the HAdV-5 fibre head; Xia et al. (1994)). The CD- and IJ-loops contain 7 residues each, while the DG-loop is composed of 16 amino acid residues. Two small beta-strands, E and F, can be observed on DG-loop in the case of HAdV5 (Xia et al., 1994). Here, the-DG loop does not contain these strands, but a prominent alpha-helix (Figure 13A).

4.1.4 Structural homologues of SnAdV-1 fibre head

As mentioned, the fibre head of SnAdV-1 has little sequence homology with other adenovirus fibre heads (between 12 and 18% sequence identity). However, analogous proteins from other adenoviruses as well as from bacteriophages were identified to share the same beta-sandwich topology. A structural homology search showed receptor binding protein of bacteriophages TP901-1 (PDB code: 2F0C; Z-score=7.5; Figure 14A) and P2 (PDB code: 2BSE; Z-score=7.3) as the closest match with Z-values of more than 7. According to Holm et al. (Holm and Rosenstrom, 2010), significant similarities have a Z-score above 2 and they usually correspond to similar folds. Strong matches have a Z-score above n/10 – 4, where n is the number of residues in the query structure, giving a Z-value of around 6 in this particular case. Further hits included the avian reovirus attachment protein sigma C (PDB code: 2BT7; Z-score=5.8; Figure 14B), human adenovirus 19p fibre head domain (HAdV-19p; PDB code: 1UXB; Z-score=5.3) and human adenovirus 37 fibre head domain (HAdV-37; PDB code: 1UXA; Z-score=5.2; Figure 14C) in descending order of similarity.

The 99- and 98-residue C-terminal receptor binding domains of TP901-1 and P2 bacteriophages are beta-barrels made up of six anti-parallel beta-strands in case of the former and seven in case of the latter. Their compact structures are pretty comparable in dimensions and show little structural deviations when superposed onto SnAdV-1 fibre head (Figure 14A). A putative binding site in TP901-1 receptor binding protein has been described (Spinelli et al., 2006b). Two of the constituting residues of this site are located on the 5th strand of the beta-barrel, which is structurally equivalent to the H strand of SnAdV-1 fibre head. Aligning these strands gives an r.m.s.d. value of 0.8 Å for 6 Cα atoms. All these viruses contain DNA as their genetic material and similarities observed in their host-receptor engaging proteins points towards a possible divergent evolution, although nothing can be said with certainity.

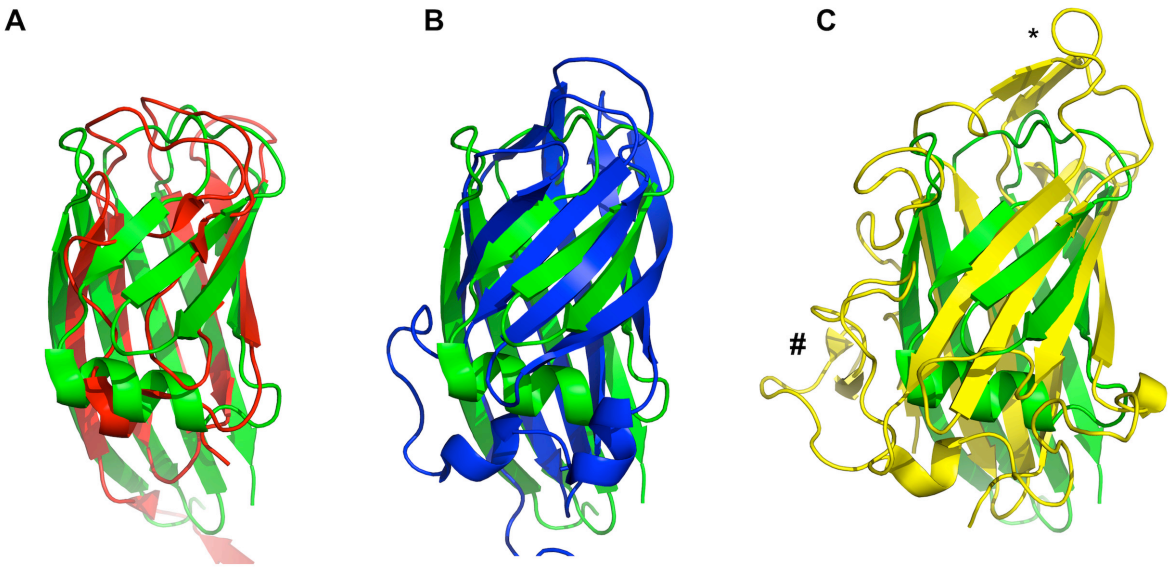


Figure 14: Structural comparison of SnAdV-1 fibre head with other viral receptor-binding proteins. Superposition of the SnAdV-1 fibre head monomer structure (green) onto the bacteriophage TP901-1 receptor-binding protein monomer (A; PDB entry 2F0C; red), onto the avian reovirus fibre head domain monomer (B; PDB entry 2BT7; blue) and the HAdV-37 fibre head monomer (C; PDB entry 2WBW; yellow). The SnAdV-1 fibre head monomer is in the same orientation as in Figure 13A. The top of the trimer is indicated with * and the side with #.

Among adenoviruses, the fibre head domains of HAdV-19p and HAdV-37 have shown most resemblance to SnAdV-1 fibre head (Figure 14C). All of these structures have a similar beta-sandwich arrangement of two beta-sheets but, unlike SnAdV-1 fibre head, structures of the formers consists of longer loops connecting beta strands. In case of human adenoviruses, loops have been implicated in fibre-receptor interactions (Bewley et al., 1999; Persson et al., 2010) and it can be speculated that a change in their length and orientation could lead to a differential binding pattern. Besides loop lengths, the average number of residues per strand is also higher in case of HAdV-19p and HAdV-37 fibre heads, which makes them taller (10 against 8 in case of SnAdV-1).

A structure based maximum likelihood phylogenetic tree constructed with previously identified homologues is shown in Figure 15. It clearly demonstrates the grouping of SnAdV-1 fibre head together with sigma C as well its closeness with bacteriophage proteins than to the adenovirus fibre heads.

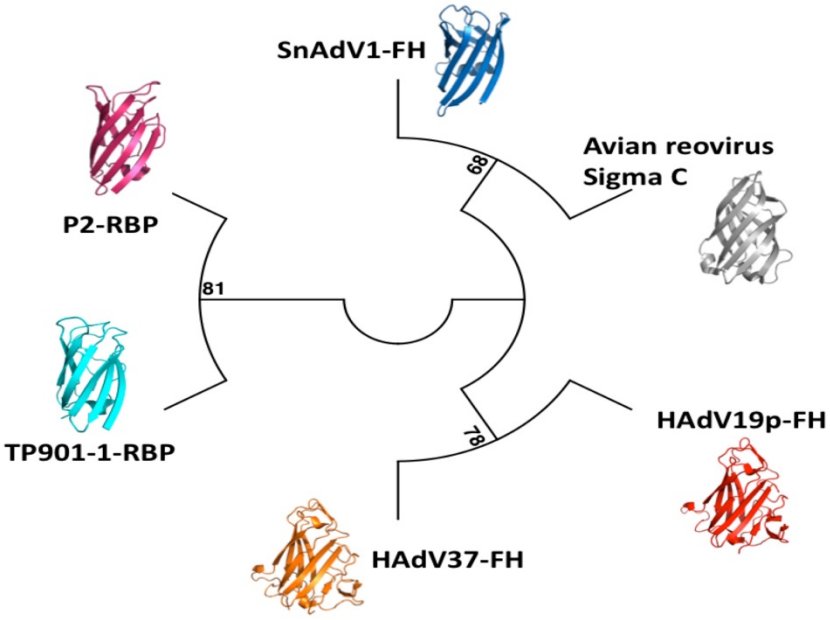


Figure 15: A structure based maximum likelihood phylogenetic tree constructed with structural homologues of SnAdV-1 fibre head. Bootstrapping employing 1000 replicates accessed correctness of the tree. Frequencies are indicated over the branch nodes.

4.1.5 Trimer interface and stability

The trimeric packing of the SnAdV-1 fibre head is highly stable and, as stated above, on a SDS gel monomeric units are only visible after heating the protein at 95 $^{\circ}$C. Each monomer has a total solvent accessible surface area of 5.9 x 10^3 Å^2, out of which 1.4 x 10^3 Å^2 (24%) gets buried upon trimer formation. Inter-monomeric contacts between assembled monomers are extensive, involving a total of 23 interfacing residues from each monomer. Eleven hydrogen bonds and ten salt bridges are contributed by each monomer in the assembly. The majority of these residues are located on strands B, C and J with some on interfacing loops. Due to this strong bonding at the trimeric core, the monomers are held tightly to each other.

4.1.6 Receptor binding

Surface exposed loops and charge distribution over the surface play crucial roles in adenovirus fibre-receptor interaction (Bewley et al., 1999; Persson et al., 2010). Where binding to receptors such as CAR and CD46 depends on the shape and orientation of loops, interaction with charged ligands takes place in a charge-dependent manner with less conformational dependency (Bewley et al., 1999; Arnberg et al., 2002; Persson et al., 2007).

As mentioned above, the loops of the SnAdV-1 fibre head are relatively short and less likely to exhibit high conformational flexibility.

Positively charged patches are naturally inclined to bind negatively charged molecules such as sialic acid, which is a common terminally attached glycan found on cell surface proteins. The surface of the SnAdV-1 fibre head is mixed in its charge distribution; the top and bottom of the head domain harbour positive patches, while the sides have a mixed charge distribution (Figure 16). Its close structural resemblance to HAdV-19p and HAdV-37 fibre heads, where sialic acid binds to the top of the trimer in a charge dependent manner (Burmeister et al., 2004), tempted me to speculate a similar saccharidic affinity behaviour for SnAdV-1 fibre head protein. In order to identify such possible ligands neoglycoconjugate microarray (NGC) experiments were carried out by Dr. Michelle Kilcoyne from the group of Prof. Lokesh Joshi at the National University of Ireland, Galway.

The result showed several glycans displaying interactions (Figure 17). Alpha-linked sialated sugars 2,3'- and 2,6'-sialyllactose, lacto-N-tetraose and xylose were found to be the most prominent ones (Figure 17). The sialated glycans and xylose were selected for further experiments. As both alpha-linked sialated sugars contain lactose moieties, it was probable that the signals were arising due to protein interacting with lactose and not with the sialic acid unit of the glycan. Experiments were repeated in presence of lactose and no binding inhibition was observed, verifying that the interactions were specific to the sialic acid unit.

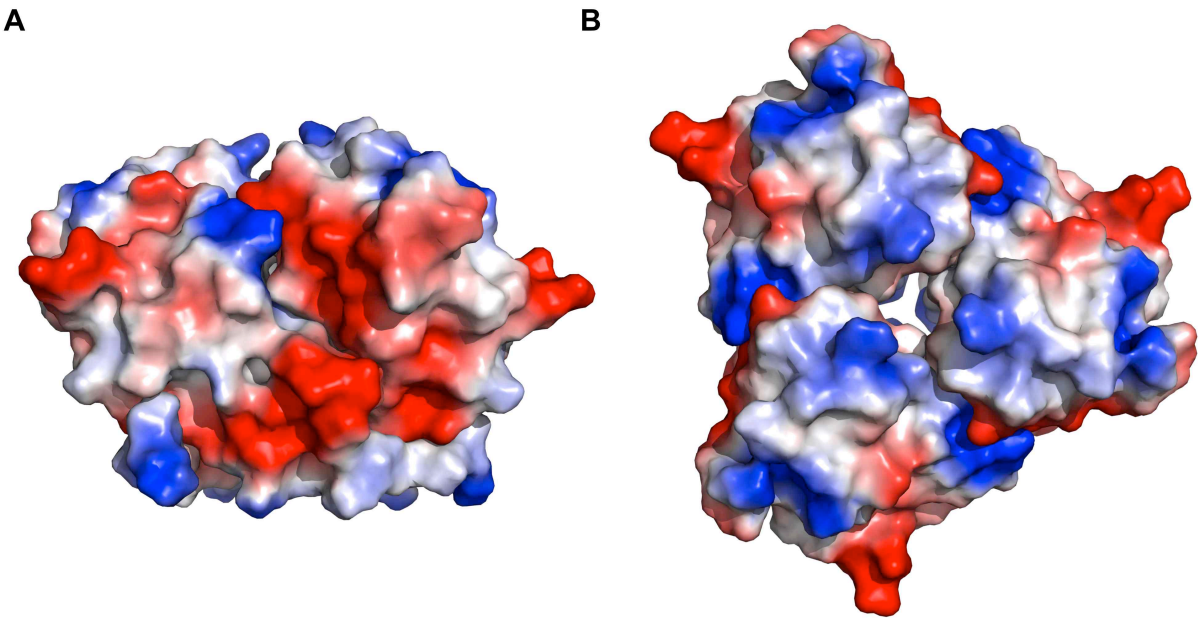


Figure 16: Qualitative electrostatic surface diagram of the SnAdV-1 fibre head. (A) Side view in the same orientation as Figure 13C. (B) Top view in the same orientation as Figure 13D.

These findings were followed up and solution experiments were performed to varify the results. STD-NMR (by Alvaro Berbís, CIB-CSIC, Madrid) and ITC (by Dr. Margarita Menéndez, IQFR-CSIC, Madrid) experiments performed with 2,3'- and 2,6'-sialyllactose confirmed the interaction, however, the affinity between the protein and the ligands was found to be very low. Co-crystallization and ligand soaking experiments did not yield ligand-bound structures, perhaps due to low binding affinity and low predicted pI of the protein.

The observed disagreement between glycan screening and solution based experiments can been seen in the light of inherent experimental differences between them. The way glycans are attached and presented on a microarray slide could significantly affects their interaction with proteins (Kilcoyne et al., 2012). Thus, its likely that one would get positive binding with relatively good strength in a glcyan array experiment, however, the same might not be obtainable to the same extent under more relaxed solution environment. I speculate that the same reason is behind these fluctuating results. Nevertheless, these observations could not completely rule out the possibilities of a glycan being the natural receptor for SnAdV-1 fibre head as the tested sugars were simpler in nature and that the virus might need a more complex partner for actual binding. Future research in this direction may give us some more clues.

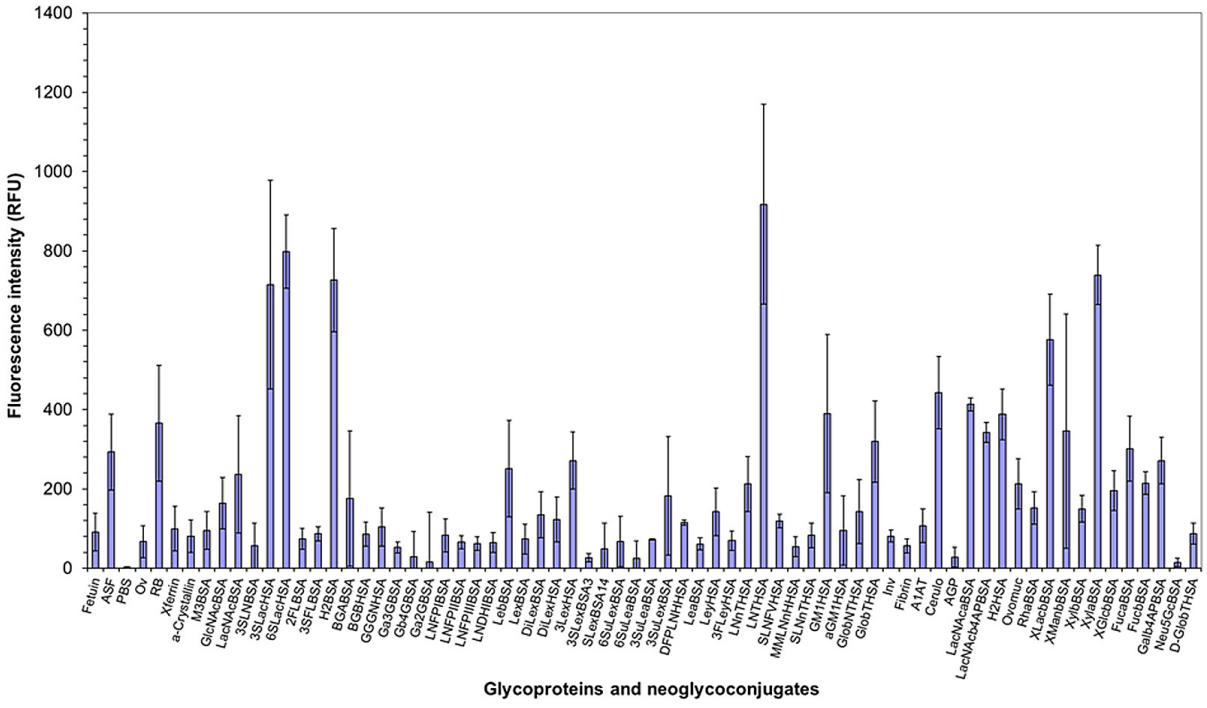


Figure 17: Glycan array of his-tagged SnAdV-1 fibre head construct 234-345. Histograms represent the average of three replicate experiments and error bars are one standard deviation. Each experiment is the median of 6 data points (6 features, 18 data points total). (See table AI in annex for details of neoglycoconjugtes and glycoproteins)**.**

4.1.7 Conclusion

The crystallographic structure of the fibre head domain of snake adenovirus 1 is described here. This is the first Atadenovirus for which the fibre head structure has been solved. The compact trimeric structure is pretty stable as evident by denaturing gel electrophoresis and has extensive non-covalent interactions among its monomeric units. From a dimension and length point of view, together with the LAdV-2 fibre 1 head, the structure is the smallest known adenovirus fibre head structure; however, it retains the same ABCJ-GHID beta-sandwich topology. It's interesting to observe that the structure is more similar to receptor binding proteins of bacteriophages than to other adenovirus fibre heads. Glycan array screening identified sialyl-lactose as a potential ligand. This suggests that the virus may utilize a cell surface glycan as its attachment receptor, however, perhaps more complex then the glycans screened here.

An important application of adenoviruses is their use as vectors for human gene and cancer therapy and for vaccination purposes. Animal adenoviruses may be of special interest due to the lack of neutralizing antibodies in human sera against them. The detailed structure of the SnAdV-1 fibre head and other animal adenovirus fibre heads, together with the identification of their natural receptors, may lead to the development of new strategies to target adenovirus vectors to cells of interest.

4.2 Structure and sialyllactose binding of the carboxy-terminal head domain of the fibre from a Siadenovirus, turkey adenovirus 3.

Turkey adenovirus 3 (TAdV-3), also known as turkey hemorrhagic enteritis virus (THEV), belongs to the *Siadenovirus* genus. The carboxy-terminal region of its fibre does not have significant sequence similarity to any other adenovirus fibre heads of known structure. The crystal structure of the trimeric fibre head domain was determined at 2.2 Å resolution. Each monomer contains a beta-sandwich, which, surprisingly, is more similar to the one in the reovirus fibre head than to other adenovirus fibres, although the ABCJ-GHID topology is conserved in all. A beta-hairpin insertion of the C-strand of each trimer subunit embraces its neighbouring monomer. The calculated electropositive surface charge of the trimer suggested that it may bind an acidic receptor, and glycan array experiments showed interaction with 2,3'- and 2,6'-sialyllactose. Surface plasmon resonance assays, nuclear magnetic resonance (NMR) spectroscopy and isothermal titration calorimetry confirmed these interactions. Kd values calculated by ITC showed that virulent fibre head has a slightly lower affinity than avirulent for 2,3'-sialyllactose. I obtained co-crystal structures with 2,3'- and 2,6'-sialyllactose and the site of interaction on 2,3'- and 2,6'-sialyllactose was mapped by saturation transfer difference NMR spectroscopy. Isothermal titration calorimetry was performed with different site-directed mutants to confirm the interaction site on the fibre head. Two sequence differences between virulent and avirulent TAdV-3 map on the fibre head; where virulent TAdV-3 contains Ile354 and Thr376, avirulent TAdV-3 contains Met354 and Met376. Although receptor-binding properties of the avirulent fibre head domain are similar to those of the virulent fibre head, we obtained evidence for differences in biophysical properties, suggesting increased flexibility of the beta-hairpin for the virulent form.

4.2.1 Purification and crystallization

4.2.1.1 Construction of expression vectors

Protein sequence analysis of the TAdV-3 fibre suggests that, like other adenovirus fibres, it contains a short, highly charged, penton base attachment sequence and a shaft domain containing putative triple beta-spiral repeats (Figure 18; van Raaij et al. (1999b)). The remaining carboxy-terminal sequence (amino acids 304–454) may form a head domain, although it shows only between 10 and 20% sequence identity when aligned with known adenovirus head domain structures. The highest sequence identity, 31 identical residues out

of 152, is with canine adenovirus 2 fibre head (Seiradake et al., 2006).

A

1-*MATPGKRSAE EPDQQTLKKS KQSDQSQGLN LAYPFDKITE FEATPPFIHV-* 50
51-GQgLDISDLS LNMRIGKgLK FENGNLVVSD QQYNVTPpLI ADQSTLGLKY-100
101-NPDvLSLTHS GALTLPTIQH pLQASAGKFE LALSSgLKSD DQGLTLDLDp-150
151-VFSTESSKFL LNCSLpLDKN SDKLTLKFGN gLGLNNDQLE NTMTYNLPLK-200
201-RDgTNVSLSF GTNFKILNEM LDLNLVApMS NSAGGLALQF KSpLSADDGI-250
251-LSIKTDTsLG ITGNKLGIRL APNSgLQITP NGLAVSVNaV QILSSPLITA-300
301-ASI**GPPTTMV TGTVSPGRAT NGQFVTKTAK VLRYKFVRWD ALLIIQFIDN**-350
351-**IGVMENPTFY RNKSIELRSA DFLSPMLNNT YIVPLNGGVR VESPTIPVQL**-400
401-**EVILENNSSF IQVGFVRLTV KNGNPHMIIQ CNPVPGNIKM IKIKSVMLFT**-450
451-**CLIG**

B

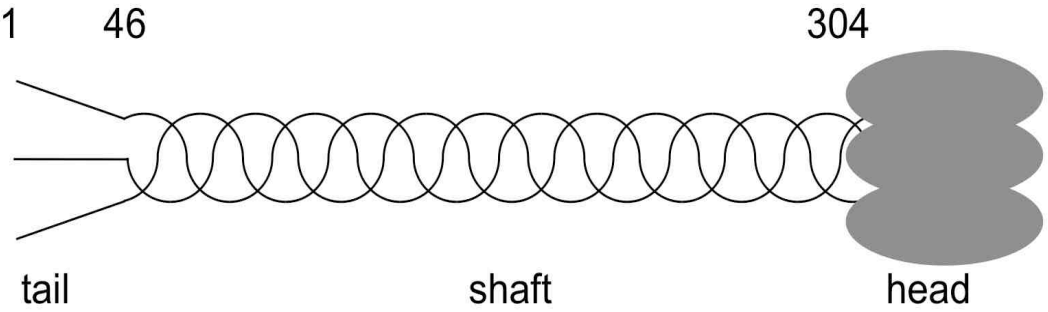


Figure 18: Predicted domain organization of turkey adenovirus 3 (TAdV-3). (A) The amino-acid sequence in one-letter code. The predicted virus-binding tail (residues 1-45) is shown in italics, the shaft domain (46-303) in normal type and the expressed predicted head domain (304-454) in bold. Putative triple beta-spiral repeats in the shaft domain are indicated by underlining the mainly hydrophobic residues that may form the central hydrophobic core and peripheral hydrophobic patches (van Raaij *et al.*, 1999b) and putting the turn amino acid (generally Gly or Pro) of each potential triple beta-spiral repeat in small type. (B) Schematic drawing of the fibre indicating the tail, shaft and head domains.

At the protein level, the major differences found between the virulent and avirulent variants of the virus are two amino acid mutations in the fibre protein: Ile354 and Thr376 in the virulent form are replaced by methionines in the avirulent form (Beach et al., 2009a). Hence, in order to generate the virulent fibre head, a two-step site directed mutagenesis approach was adopted where aforementioned mutations were generated.

4.2.1.2 Protein purification and crystallization

Expression vectors for native and mutant TAdV-3 fibre proteins were used to produce milligram amounts of soluble protein. Proteins were eluted from the nickel column at relatively high imidazole concentrations, consistent with them being stable trimers as expected; and thus containing three six-histidine tags per protein unit. In cation-exchange chromatography, the proteins (predicted pI 9.9) eluted in several peaks around 0.3–0.5 M sodium chloride when a linear gradient was used, each of which contained pure or nearly pure fibre head protein (Singh et al., 2013a).

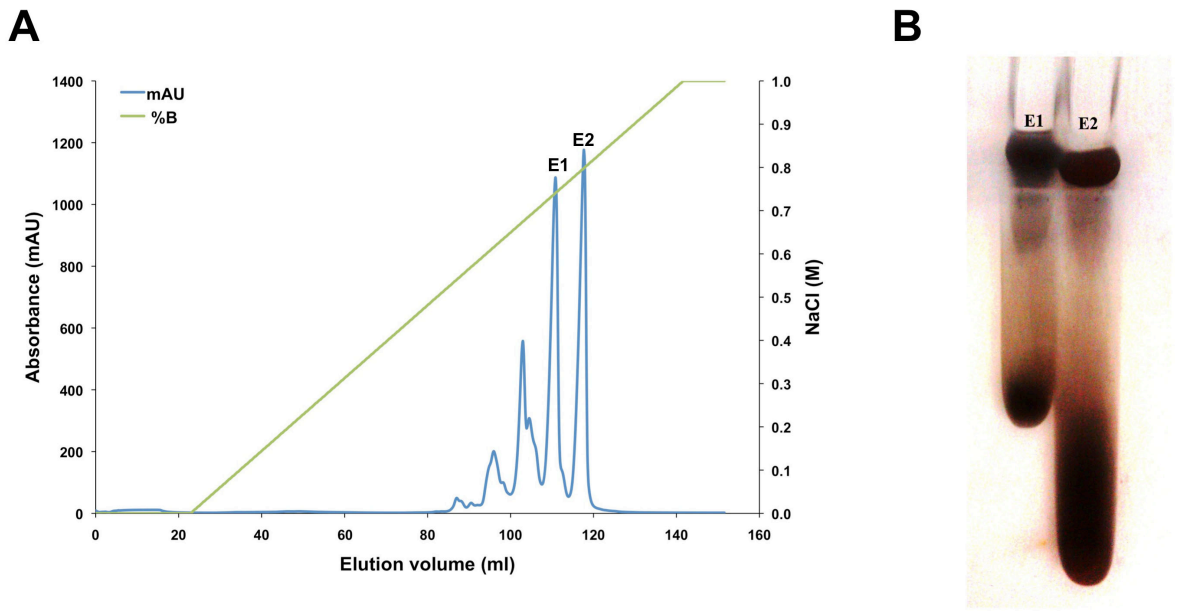


Figure 19: Purification of TAdV-3 avirulent fibre head. (A) Chromatogram of the final strong anion exchange purification step. Pure protein elutes in several peaks; protein from each peak was kept separate. Protein from the highest two peaks (E1 and E2) was used for crystallization trial. In green the sodium chloride concentration gradient is indicated. (B) Native PAGE-gel run for protein eluted in peaks E1 and E2 showing differential mobility. Gel is stained with silver stain.

Figure 19A shows the cation-exchange chromatography profile for avirulent fibre head. Protein from the different peaks was kept separate and, after concentration, used for crystallization. In all cases, isomorphous crystals for protein from each of two major peaks were obtained in conditions containing diammonium phosphate as a precipitant, sodium chloride as an additive and sodium citrate at pH 5.5-5.8 as a buffer. It appears that the protein exists in several states differing in charge, but it is not clear to us what causes these differences. A native PAGE gel shows the differential mobility of proteins from two peaks (Figure 19B). It's noteworthy that the same was observed for HAdV-2 fibre head expressed

in baculovirus-infected insect cells (van Raaij et al., 1999a). Both the fibre variants as well as the mutants showed almost identical elution profiles during purifications.

A total of about 500 conditions were tried and crystals were obtained within a time period of four days in at least three of those conditions, all containing diammonium phosphate:

1. 1.0 M diammonium phosphate, 0.1 M imidazole pH 8.0),
2. 1.0 M diammonium phosphate, 0.1 M Tris pH 8.5)
3. 1.0 M diammonium phosphate, 0.2 M sodium chloride, 0.1 M sodium citrate pH 5.5).

Crystals for his-tagged avirulent and virulent fibre heads were obtained (Figure 20A). The crystals were cubic in shape and belonged to space group *I*23 (unit-cell parameter 99 Å). Complete and highly redundant X-ray diffraction data sets were collected from each of the native crystal and a avirulent fibre head crystal derivatised with selenomethionine, the latter using a wavelength at which the anomalous signal of the Se atoms was maximized (Table 13; Figure 20B). The data used extended to 2.0 Å and 2.3 Å in case of avirulent and virulent fibre native data sets respectively and 2.14 Å resolution for the derivative crystal. Merging R factors for shells higher than 2.4 Å resolution were quite high; however, the correlation between independent measurements of the same reflection and mean I/"(I) in the highest resolution bins are still significant (Table 12)

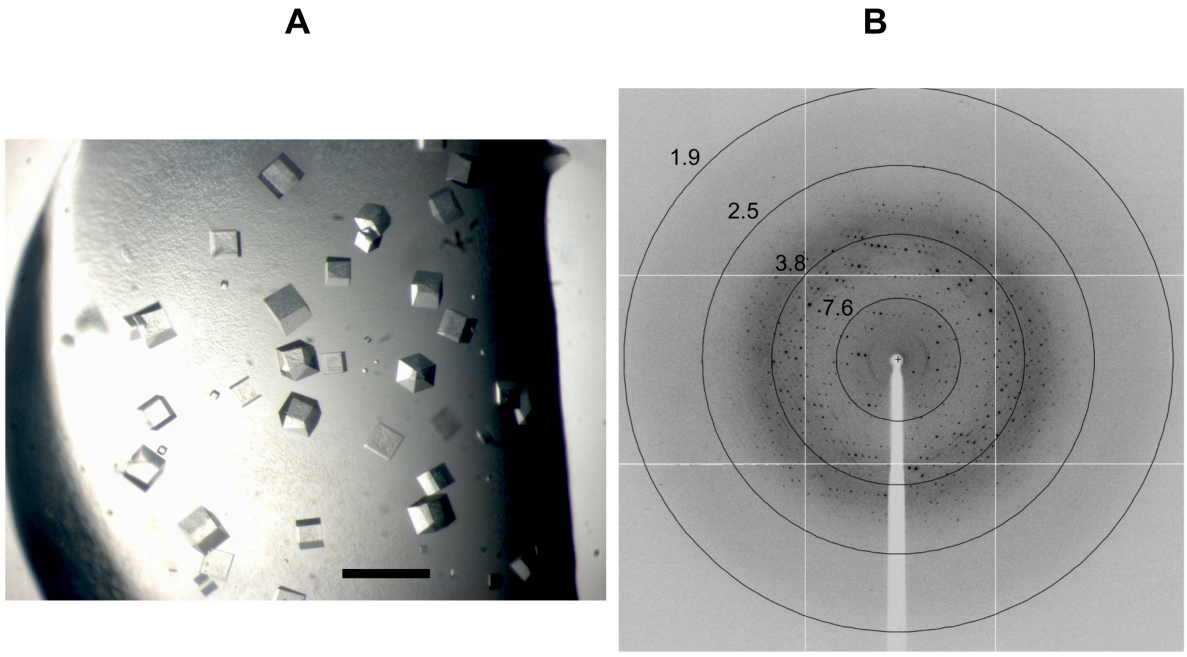


Figure 20: Crystallization and diffraction analysis of TAdV-3 fibre head. (A) Crystals belonging to space group *I*23. The scale bar measures approximately 0.5 mm. (B) Diffraction pattern of TAdV-3 fibre head crystals. The beam centre is indicated by a crosshair and resolution rings (in Å) are shown.

Table 12: Data quality as a function of resolution (For TAdV-3 fibre data sets)

Native (avirulent)					
Resolution (Å)	Reflections	R_{sym}	Mean <I/s(I)>	CC-Imean	CC-anom
6.96	287	0.029	36.0	0.999	0.076
4.92	490	0.044	36.4	0.999	0.341
4.02	642	0.079	35.0	0.996	-0.236
3.48	725	0.075	31.1	0.997	-0.021
3.11	833	0.074	26.6	0.998	-0.015
2.84	906	0.104	19.0	0.997	-0.006
2.63	987	0.182	12.7	0.990	-0.080
2.46	1067	0.313	8.1	0.974	-0.028
2.32	1136	0.479	5.4	0.928	-0.029
2.20	1183	0.694	3.8	0.856	0.019
Native (virulent)					
Resolution (Å)	Reflections	R_{sym}	Mean <I/s(I)>	CC-Imean	CC-anom
7.27	256	0.029	35.0	0.999	0.160
5.14	440	0.046	33.0	0.999	-0.241
4.20	559	0.061	35.0	0.997	-0.081
3.64	658	0.060	30.9	0.997	-0.267
3.25	722	0.069	25.9	0.997	-0.082
2.97	821	0.087	19.8	0.996	-0.205
2.75	869	0.126	12.7	0.993	-0.136
2.57	948	0.202	8.1	0.982	-0.059
2.42	998	0.365	5.0	0.937	-0.015
2.30	1053	0.534	3.6	0.903	0.081
Selenomethionine					
Resolution	Reflections	R_{sym}	Mean <I/s(I)>	CC-Imean	CC-anom
6.96	287	0.049	59.3	1.000	0.916
4.92	498	0.083	54.5	0.999	0.689
4.02	634	0.093	59.3	0.999	0.430
3.48	732	0.074	51.1	0.999	0.644
3.11	844	0.083	44.0	0.999	0.748
2.84	916	0.128	31.8	0.999	0.616
2.63	999	0.206	21.8	0.997	0.537
2.46	1066	0.350	14.1	0.993	0.314
2.32	1121	0.538	9.6	0.985	0.131
2.20	1201	0.814	6.5	0.960	0.080

4.2.2 Structure determination of the TAdV-3 fibre head domain

Out of six selenium sites expected from the sequence, four were located by the automated phasing procedure, and refinement of these sites resulted in high-quality phases (see phasing statistics in Table 13). Solvent flattening resulted in an easily interpretable map in which 137 amino acids were automatically traced. The asymmetric unit contains one protein monomer and the biologically relevant trimer is generated by the crystallographic three-fold symmetry axis. Manual completion resulted in a model consisting of 138 amino acid residues (317-454), one phosphate and 47 water molecules. The model was also independently refined against native data collected from crystals of the avirulent form and virulent forms, and here manual adjustment resulted in a model containing the same 138 residues plus ordered solvent (Table 13). The phosphate ion mediates a crystal contact between Lys421 and Asn422 of one monomer and Arg368 and Lys439 of a monomer belonging to another trimer. No density was visible for residues N-terminal to 317 or for the purification tag.

The final refined structures have good geometry and acceptable Ramachandran statistics (Table 13). The structures each contain one genuine cis-peptide bond, between Ser374 and Pro375. The head domain appears to start at around residue 317, rather than around 304 as we predicted. It is possible that the shaft domain ends around residue 303, and the head domain, starting at residue 317, is connected to the shaft domain by a short linker sequence, which also occurs in other virus fibres (for example HAdV-2; van Raaij et al. (1999a)).

Table 13: Crystallographic data measurement and refinement statistics for native and selenomethionine derivative TAdV-3 fibre head crystals

	Avirulent native	Selenomethionine	Virulent native
Data collection			
Synchrotron and beamline	ESRF ID29	ESRF ID29	ALBA BL13-XALOC
Wavelength (Å)	0.9768	0.9791	0.9795
Crystal-detector distance (mm)	263.0	329.6	381
Oscillation range (°)	1.0	1.0	0.5
Number of images	100	360	180
Space group	*I*23	*I*23	*I*23
Cell edge (Å, a=b=c)	98.57	98.75	98.96
Mosaicity (°)	0.5-0.6	0.2	0.3-0.5
Resolution range (Å)	30-2.2 (2.32-2.20)*	30-2.2 (2.32-2.20)*	31.3-2.3 (2.42-2.30)*
Reflections	8256 (1183)	8297 (1201)	7327 (1053)
Multiplicity	11.6 (12.0)	43.2 (43.1)	8.9 (9.2)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
Mean <I/s(I)>	17.5 (3.8)	29.0 (6.5)	6.3 (1.4)
R_{sym} (%)†	8.8 (69.4)	10.4 (81.4)	6.2 (53.4)
V_M ($Å^3Da^{-1}$)‡	1.94	1.94	2.02
Wilson B ($Å^2$)	43.1	45.1	56.8
Phasing			
Heavy atom sites		4 Se	
Correlation coefficient (all/weak)		55.47 / 33.95	
Patterson figure of merit		22.19	
Correlation coefficient (E)		0.500	
R-cullis** (anomalous, acentric)		0.574	
Phasing power** (anomalous differences)		2.066	
FOM = cos(phase error)** (acentric/centric)		0.4152 / 0.0728	
Solvent flattening (44.6% solvent content)			
R-factor (before/after)		0.4800 / 0.2760	
Overall corr. on $\|E\|^2$ (before/after)		0.3057 / 0.7166	
Correlation on $\|E\|^2$ / contrast (original/inverted)		0.4630 / 0.2240	
Refinement			
Resolution range (Å)	28.51-2.20 (2.3-2.2)	28.51-2.20 (2.3-2.2)	30.0-2.30 (2.4-2.3)
Reflections used	7866 (1128)	7909 (1145)	6962 (984)
Reflections used for R_{free}	384 (55)	385 (56)	334 (40)
R-factor	0.199 (0.251)	0.207 (0.254)	0.197 (0.245)
R-free	0.242 (0.332)	0.252 (0.295)	0.239 (0.321)
Protein/phosphate/water atoms	1092/5/57	1092/5/47	1091/10/40
<B> overall/protein/ion/water ($Å^2$)	42.0/41.7/83.8/43.7	42.1/41.9/74.9/43.3	64.7/65.4/59.3/52.8
Ramachandran statistics(%)	97.1/99.3	96.3/99.3	94.1/99.3
r.m.s.d. bonds (Å)/angles (°)	0.014/1.5	0.014/1.5	0.013/1.5
PDB code	3ZPE	3ZPF	4CW8

* Values in parentheses are for the highest resolution bin (where applicable).

† $R_{symm} = \Sigma_h\Sigma_i|I_{hi}-\langle I_h\rangle|/\Sigma_h\Sigma_i|I_{hi}|$, where I_{hi} is the intensity of the *i*th measurement of the same reflection and $\langle I_h\rangle$ is the mean observed intensity for that reflection.

‡ According to Matthews (1968)

4.2.3 Structure of the TAdV-3 fibre head

The TAdV-3 fibre head is around 5 nm high and about 4.5 nm in diameter (Figure 21A). SnAdV-1 fibre head is 3.8 nm high and 4.6 nm in diameter while HAdV-5 fibre head is around 4 nm high and 6.2 nm in diameter. Thus, a higher height-to-diameter ratio of TAdV-3 fibre head makes it appear more slender and cylindrical (Figure 21A).

The topology of TAdV-3 fibre head is similar to that reported for other adenovirus fibre head structures (Xia et al., 1994), despite the low sequence homology (only 10-20% sequence identity when aligned with known adenovirus head domain structures).

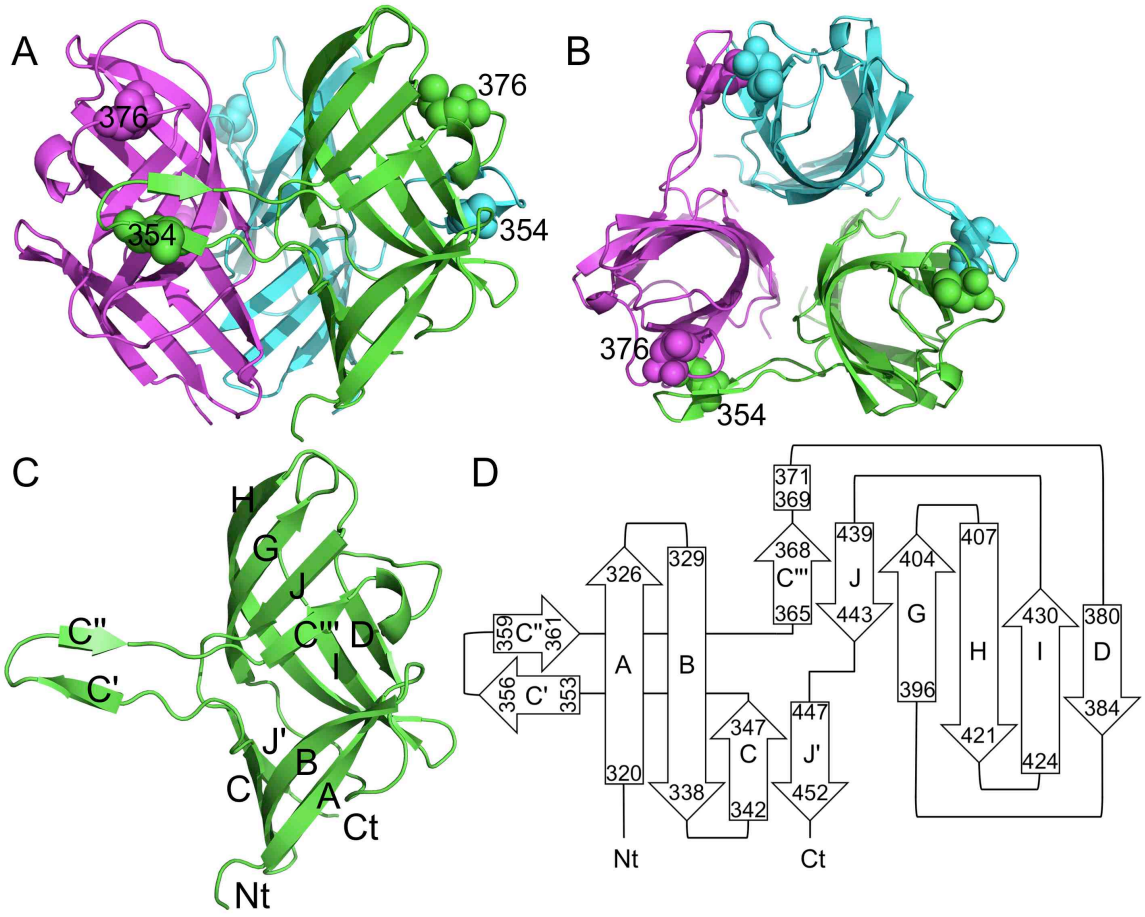


Figure 21: Structure of the TAdV-3 fibre head. Side view (A) and top view (B) of the fibre head trimer. Between panels A and B the structure has been turned approximately 90º towards the reader. Each monomer is colored differently. (C) Side view of a monomer with the N-terminus, C-terminus and beta-strands labeled. (D) Topology diagram of the monomer labelled as in C, but with the start and end residue of each strand also indicated. In panels A and B Met354 and Met376 are shown in space-filled representation and labeled.

Each monomer contains an eight-stranded beta-barrel (ABCJGHID), in which the C- and J-strands are kinked (Figure 21D). The C-strand is interrupted by a 27-residue stretch that forms a beta-hairpin "arm" contacting a neighbouring monomer (see below). The bottom

section of the C-strand was named C, the beta-strands of the arm C' and C", and the top section of the C-strand C'" (Figure 21C). Similarly, the J-strand is divided into an upper part (J) and a lower part (J'). The upper part interacts both with the C'"-strand and the G-strand, while the lower part interacts with the C-strand. Beta-strands E and F present in the HAdV-5 fibre head (Xia et al., 1994) are absent in the presented structure. The beta-strands in the fibre head are connected with predominantly short loops, in some cases simple beta-turns. The C'"D-loop contains a few residues (369-371) in 3_{10}-helical conformation. The long DG-loop runs along the bottom of the beta-sandwich and contains amino acid side chains that contribute significantly towards the hydrophobic core of the monomer. The insertion of a beta-hairpin in the C-strand contacting a neighbouring monomer is a unique feature of this structure.

4.2.4 Structural differences between the avirulent and virulent fibre head structures

TAdV-3 (also known as THEV) is found in virulent as well as in avirulent form. The virulent form is highly pathogenic, while the avirulent form can cause some clinical symptoms but seldom results in death of the infected birds. It is therefore used in vaccination of turkeys against turkey hemorrhagic enteritis. At the protein level, differences found between the two variants of the virus are two amino acid mutations in the fibre protein: Ile354 and Thr376 in the virulent form are replaced by methionines in the avirulent form (Beach et al., 2009a). The mechanism that causes the difference in virulence is unknown, but it is perhaps not a coincidence that these residues are near potential glycosylation sites. Sequence analysis suggests four potential *N*-linked glycosylation site motifs (N-X-T/S): 356-NPT-358, 362-NKS-364, 378-NNT-380 and 406-NNS-408. Any mutation in the region surrounding glycosylation sites may change the hemagglutination profile of the virus (Arnberg et al., 1997). Although Met354 and Met376 have a 36 Å intra-monomer distance, they approach each other to 13 Å inter-monomer distance in the trimer.

When the structures of avirulent and virulent forms are compared, the structures are virtually identical with an r.m.s.d. of 0.53 Å when the C-alpha atoms of residues 317-454 are superposed. The largest difference is in the C'C"-loop, which moves upwards a few Å (Figure 22). This loop is near both residues that differ between the avirulent and virulent forms, i.e. near residue 354 located in the C'-strand of the same monomer (Met in the avirulent form and Ile in the virulent form), but also near in space to residue 376 in the C"D-loop of a neighboring monomer, which is Met in the avirulent form and Thr in the virulent form. The

reason for the different position of the C'C"-loop is not clear, as the two mutated residues occupy very similar positions in their respective structures.

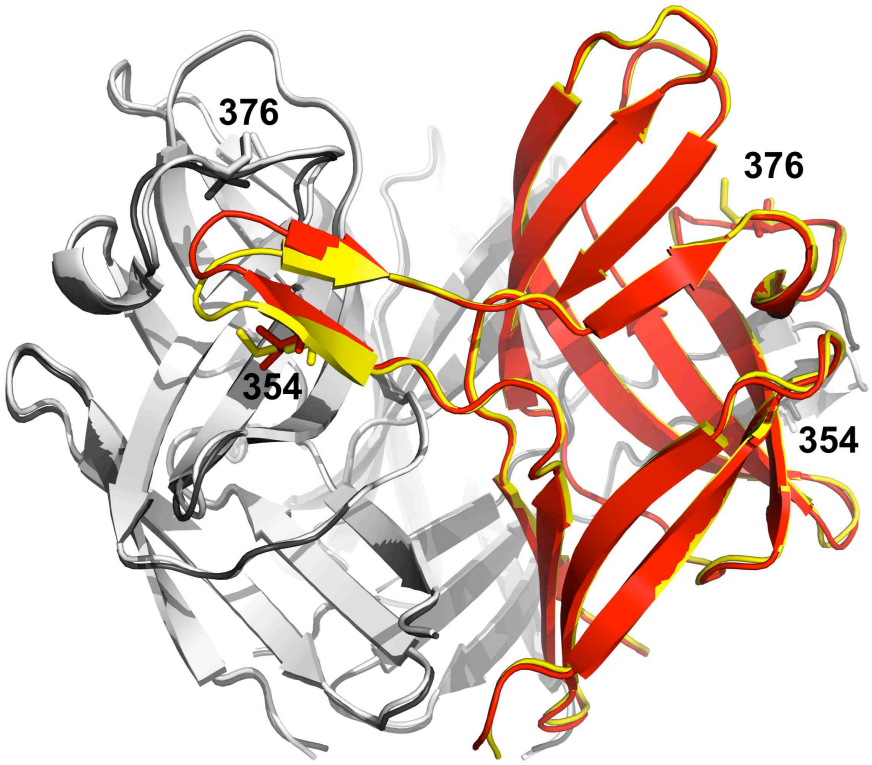


Figure 22: Comparison between the structures of the avirulent and virulent versions of the TAdV-3 fibre head trimer. One monomer of the avirulent form is shown in yellow secondary structure cartoon representation, the other two in white, while the virulent fibre head trimer is shown in red and gray, respectively. Residues 354 and 376, which differ between the two forms, are shown in stick representation and labeled.

4.2.5 Trimer stability and behaviour in solution

Each TAdV-3 fibre head monomer has a surface area of 9.4 x 10^3 Å^2, of which 2.6 x 10^3 Å^2 (27 %) is buried in the trimer. The calculated energy gain upon trimer formation is around 60 kcal/mol and a salt bridge is formed between Glu401 and Arg390 of a neighbouring monomer. These properties are comparable to other adenovirus and reovirus fibre heads, suggesting that the stability of the trimer is comparable. As mentioned above, a beta-hairpin "arm" (residues 349-364) contacts a neighbouring monomer. Out of the sixteen residues, eleven participate in inter-monomer contacts, so the arm may well provide extra trimer stability. The stable nature of trimers is evident from thermal denaturation profiles where both variants showed high thermal resistance. Both proteins have melting temperatures (Tm) close to 80 °C (Figure 23).

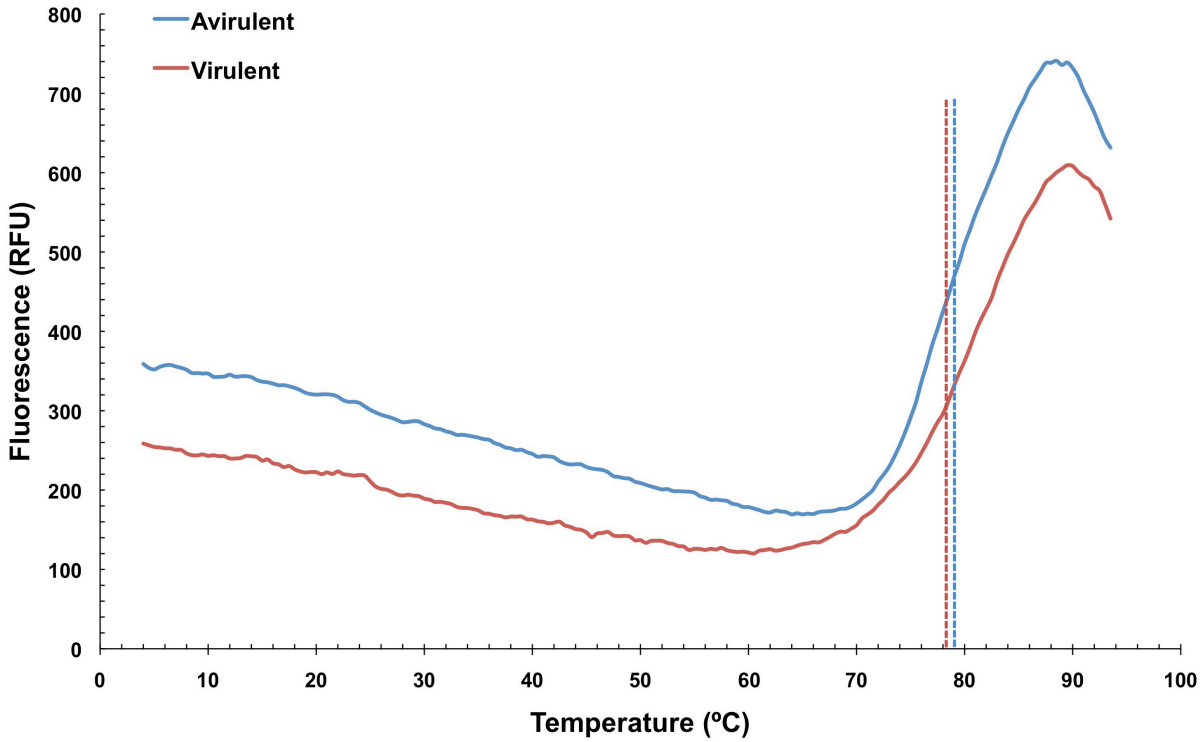


Figure 23: Thermofluor unfolding assay of avirulent and virulent TAdV-3 fibre head proteins. Curve represents average of three independent reactions for each protein carried out at same time. The mid-points of melting curves indicating Tm are shown by blue (avirulent) and brown (virulent) dotted lines.

Differences in behaviour of avirulent and virulent fibre head variants in solution were observed. Although both proteins are pretty soluble, the virulent fibre head had a tendency to aggregate. A higher inter-particle interaction could be assumed as a possible reason behind aggregation. Further evidence came from diffusion ordered spectroscopy (DOSY) NMR, which can tell about rate of molecule diffusion at a given concentration. In DOSY experiments carried out at room temperature, the self-diffusion value for virulent fibre was smaller compared to avirulent, indicating a bigger molecular size for the former (Figure 24), which possibly be due to inter-molecular interactions of virulent fibre trimers. I speculate that perhaps the beta-hairpin "arm" is less stable (conversely more flexible) in the virulent version and exists in a more open conformation, thus making the protein more prone to establish contact with other protein molecules in solution. However, structurally I have not seen evidence for this flexible behaviour and both protein variants appeared almost identical except for a slightly upward shift in the beta-hairpin arm of virulent fibre head as described above.

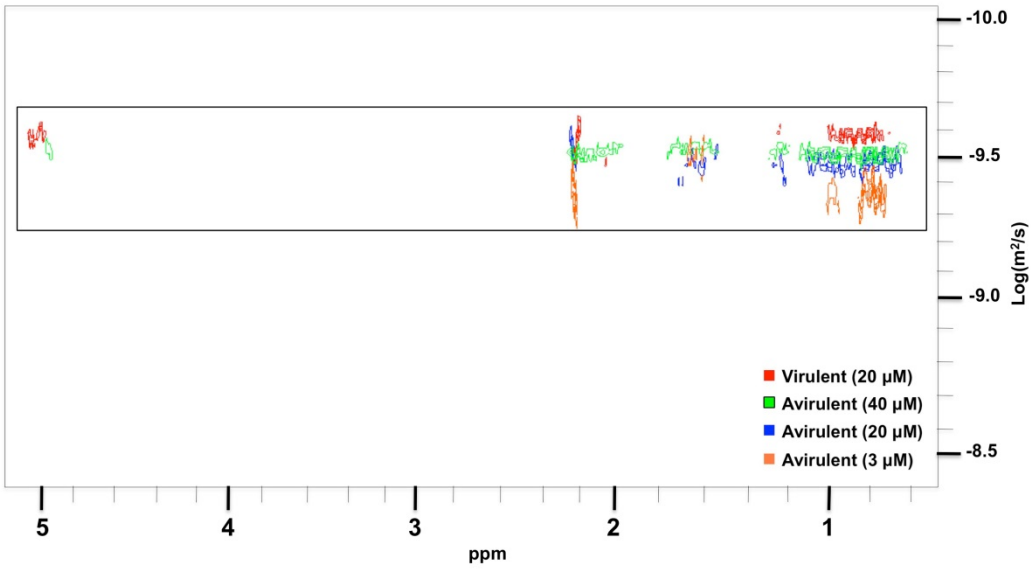


Figure 24: Behaviour of TAdV-3 fibre head proteins in solution. Diffusion ordered spectroscopy (DOSY)-NMR profiles of avirulent and virulent fibre heads at different concentrations. Y-axis denotes rate of diffusion; larger the size of the molecule, the smaller will be its diffusion rate.

Despite of the observed difference in beta-hairpin "arm" positioning between virulent and avirulent fibre heads, their near-identical unfolding temperatures indicate that the overall trimeric assembly is not getting affected by this variation. Although there is certainly a possibility that this variation may bring local changes in the structures, which in turn may have functional implications (discussed below).

4.2.6 Comparison with other fibre heads

A structure similarity search showed that the TAdV-3 fibre head structure is more similar to reovirus fibre heads sigma1 and sigmaC (Chappell et al., 2002; Guardado Calvo et al., 2005) than to other adenovirus fibre heads. When the TAdV-3 fibre head was superposed on the mammalian reovirus sigma1 fibre head domain (PDB entry 2OJ5; Figure 25A) the r.m.s.d. was 2.9 Å for 106 superimposed C-alpha atoms (Z-score 8.3). After these two proteins, the next most similar protein identified was the lactococcal bacteriophage TP901-1 receptor-binding protein (PDB entry 2F0C; Figure 25B) with an r.m.s.d. of 2.9 Å for 90 superimposed C-alpha atoms (Z-score 5.7). Finally, when the structure was superposed onto the HAdV-5 fibre head domain (PDB entry 1KNB; Figure 25C), the r.m.s.d. was 3.8 Å for 96 superimposed C-alpha atoms (Z-score 4.8).

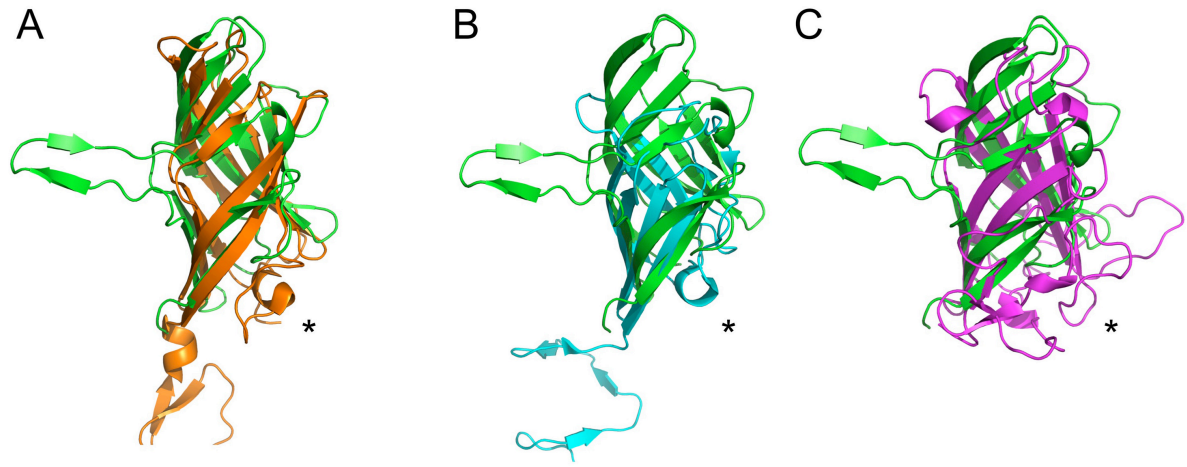


Figure 25: Structural homologues of the TAdV-3 fibre head. Superposition of the TAdV-3 fibre head monomer (green) onto mammalian reovirus sigma1 head domain (orange, panel A: PDB entry 2OJ5, r.m.s.d 2.8 Å for 104 superimposed C-alpha atoms), onto lactococcal bacteriophage TP901-1 receptor-binding protein (cyan, panel B: PDB entry 2F0C, r.m.s.d. 3.1 Å for 88 superimposed C-alpha atoms) and onto HAdV-5 head domain (magenta, panel C: PDB entry 1KNB, r.m.s.d. 2.8 Å for 83 superimposed C-alpha atoms). They are all in approximately the same orientation as in Figure 21C. Asterisks show the location of the DG-loops.

The structure shares the secondary structure topology with the reovirus fibre heads, and also the kinked nature of strands C and J (the latter is called H in the reovirus fibre heads). Other adenovirus fibre heads do not have kinks in the C-strand and only have the equivalent of the J'-strand and not the J-strand. An evolutionary link has been proposed between adenovirus fibres and reovirus attachment proteins, based on their striking similarity in beta-sheet composition and topology of the head domain and the presence of beta-spiral repeats in their shaft domains (Chappell et al., 2002). However the eight anti-parallel beta-strands, which are present in both of these proteins, have a somewhat different arrangement, even if the topology is the same. In the case of adenovirus fibres these anti-parallel sheets are arranged in beta-sandwich architecture of two four-stranded sheets, while in reovirus sigma1 and sigmaC the eight beta-strands form a beta-barrel, i.e. the two sheets contact each other at the ends (Chappell et al., 2002; Guardado Calvo et al., 2005). In the case of the TAdV-3 fibre head, the two beta-sheets contact each other only at one end, the one that is located towards the inside of the trimer (Figure 21A).

Besides reovirus proteins, the search also revealed the lactococcal phage TP901-1 receptor-binding protein (Veesler et al., 2012) as a structural homologue. Like the TAdV-3 fibre head, the C-terminal head domain of this protein (residues 63 to 163) contains an eight-stranded curved beta-sheet. However, the lactococcal phage protein head domain is much smaller, with most beta-strands being much shorter (i.e. the TAdV-3 fibre head is "higher").

When comparing adenovirus fibre heads only, the closest neighbour is that of the porcine adenovirus 4 (Guardado-Calvo et al., 2010), whereas the closest human adenovirus analogues appeared to be the fibre head of HAdV-19p and HAdV-37. However, as all other adenovirus fibre heads known, they contain a beta-sandwich of four beta-strands in each sheet, in which the two sheets do not contact each other at any of the sides to form a larger curved sheet or beta-barrel.

None of the trimeric receptor-binding domains mentioned before show the beta-hairpin insert in the C-strand. However, both reovirus and adenovirus fibre heads have a much more elaborate DG-loop than the TAdV-3 fibre head, which in both cases is involved in receptor-binding (Bewley et al., 1999; Kirchner et al., 2008). SnAdV-1 fibre head also lacks any such beta-hairpin insertion in its structure but it does have a long DG loop, albeit occupied by a prominent alpha-helix (see section 4.1.3).

4.2.7 Surface potential of the trimer and sialyllactose binding

Different types of adenovirus fibre heads have different surface charge distributions and predicted iso-electric point values. Human adenovirus 3 has a predicted iso-electric point of 5.0, while HAdV-37 has a value of 9.1. The TAdV-3 fibre head has an even higher predicted iso-electric point of 9.9 and large electropositive patches are present on the surface of the trimer (Figure 26). HAdV-37 fibre head binds sialic acid (as well as CAR; (Burmeister et al., 2004; Seiradake et al., 2009)), and the interaction with sialic acid is charge dependent (Arnberg et al., 2002). The fact that a siadenovirus-specific open reading frame exists of which the putative gene product has a high sequence similarity with bacterial sialidase proteins was also suggestive of sialic acid playing a role in Siadenovirus biology.

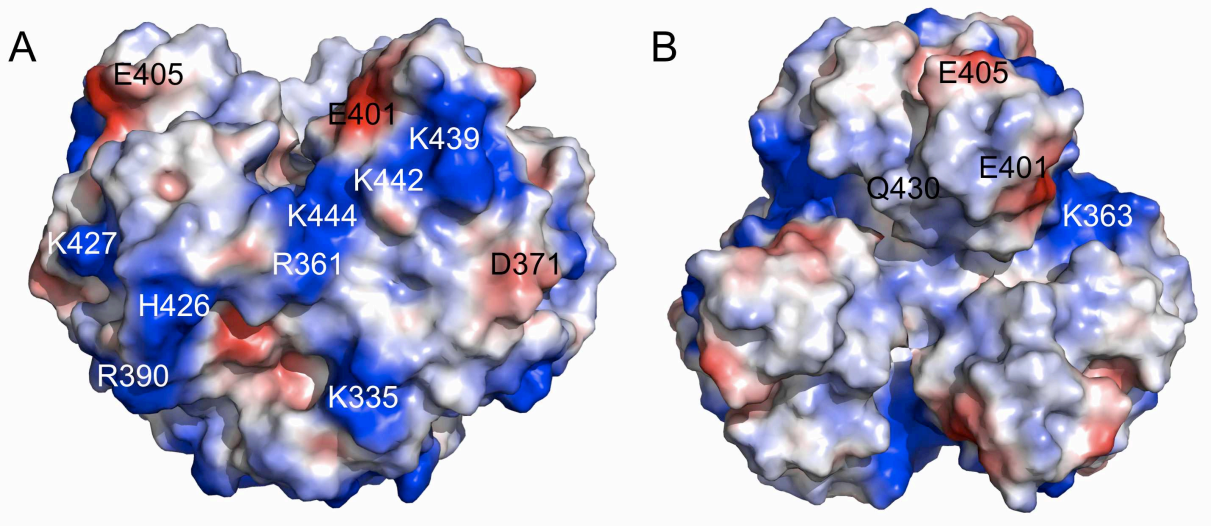


Figure 26: Surface properties and ligand binding of the TAdV-3 fibre head. Side view (A) and top view (B) with calculated negatively charged regions in red and positive regions in blue. The location of some of the polar, electropositive and electronegative amino acid side-chains that contribute to these charged regions are indicated.

Glycan microarray profiling of both variants of His-tagged TAdV-3 fibre head protein was performed with a library of 74 glycoconjugates. The proteins showed substantial interaction with two sialylated glycans, 2,3'- and 2,6'-sialyllactose, both consisting of a sialic acid moiety attached to lactose in different linkage positions (Figure 27). The observed fluorescence intensities were 5,000 and 6,100 RFU for 3'- and 6'-sialyllactose, respectively, which was significantly higher than the experimental threshold (Figure 27). No binding inhibition was observed in the presence of lactose and galactose, indicating that the sialic acid portion of the carbohydrates were involved in contact with the proteins. Xyloses linked to BSA through a 4-aminophenyl linker were also identified as a potential ligand for the viral proteins.

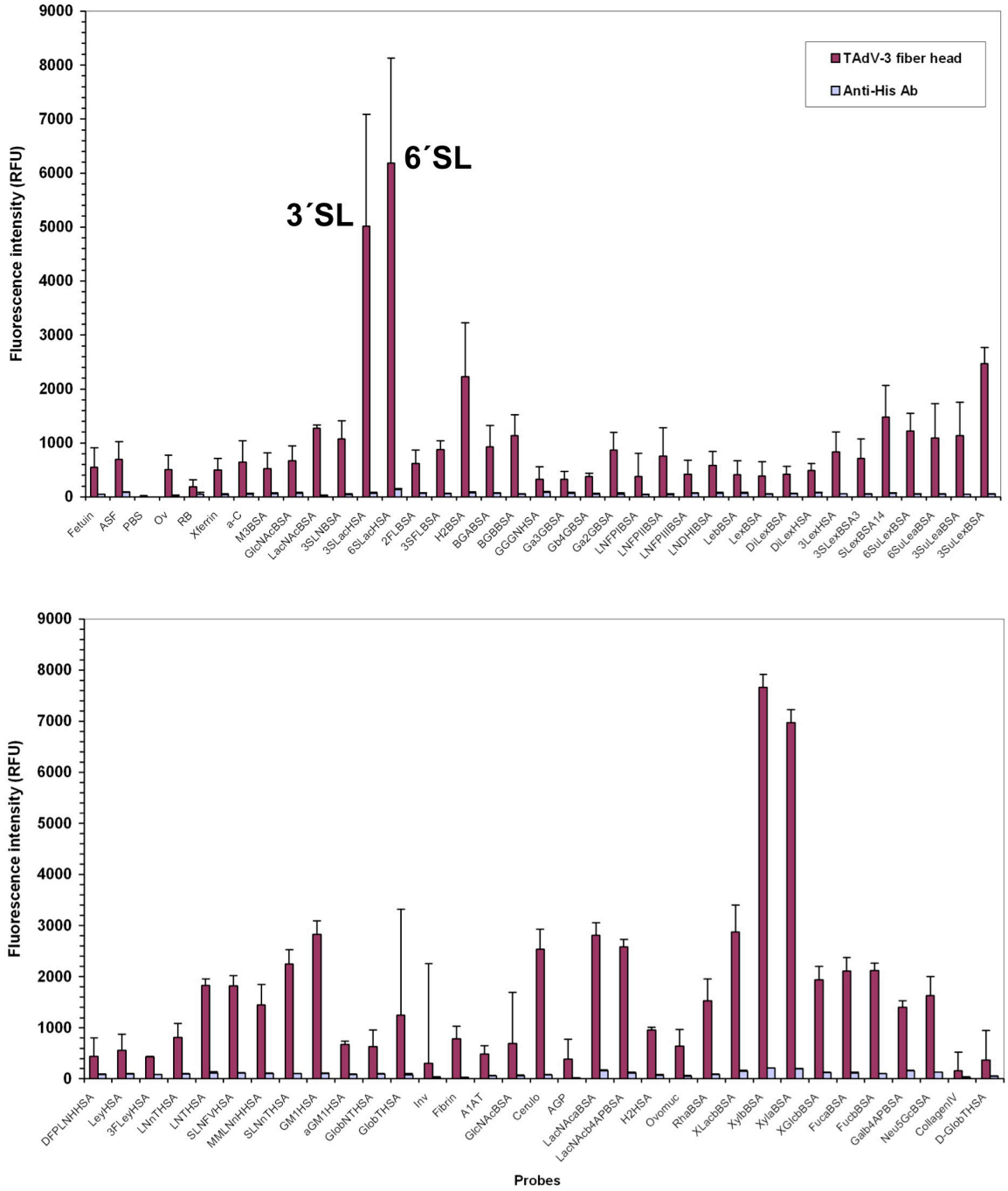


Figure 27: Histogram representing the fluorescence intensity of avirulent TAdV-3 fibre head bound to presented glycans on the microarray surface detected by fluorescently-labelled anti-his antibody. Histograms represent the average of three replicate experiments and error bars depict one standard deviation of the mean calculated over three microarray slides. Bars for 2,3'- and 2,6'-sialyllactose are indicated. (See table AI in annex for details of neoglycoconjugtes and glycoproteins).

To validate these results, TAdV-3 fibre head protein binding to 2,3'- and 2,6'-sialyllactose was examined by STD-NMR. The STD experiment is based on the transfer of magnetization between the protein, which is selectively irradiated by a train of radiofrequency pulses, and any binding ligands (Meyer and Peters, 2003). For complex ligands, such as

oligosaccharides, those atoms lying closer in space to the protein surface receive a higher amount of magnetisation, providing information on the ligand contact atoms (Mayer and Meyer, 1999). The carbohydrates 2,3'- and 2,6'-sialyllactose in the presence of either the virulent or the avirulent versions of TAdV-3 fibre head proteins resulted in STD-positive peaks, which indicated that both ligands were bound by the proteins in solution (Figure 28). In all instances, the highest degree of saturation was observed at the sialic acid, and to a much lesser degree, at the galactose region (Figure 29). These experiments not only confirmed the protein-carbohydrate interaction but also mapped the protons of the carbohydrate residues with which the proteins interact, indicating that binding of sialyllactose to TAdV-3 fibre head occurs mainly at the level of the sialic acid residue.

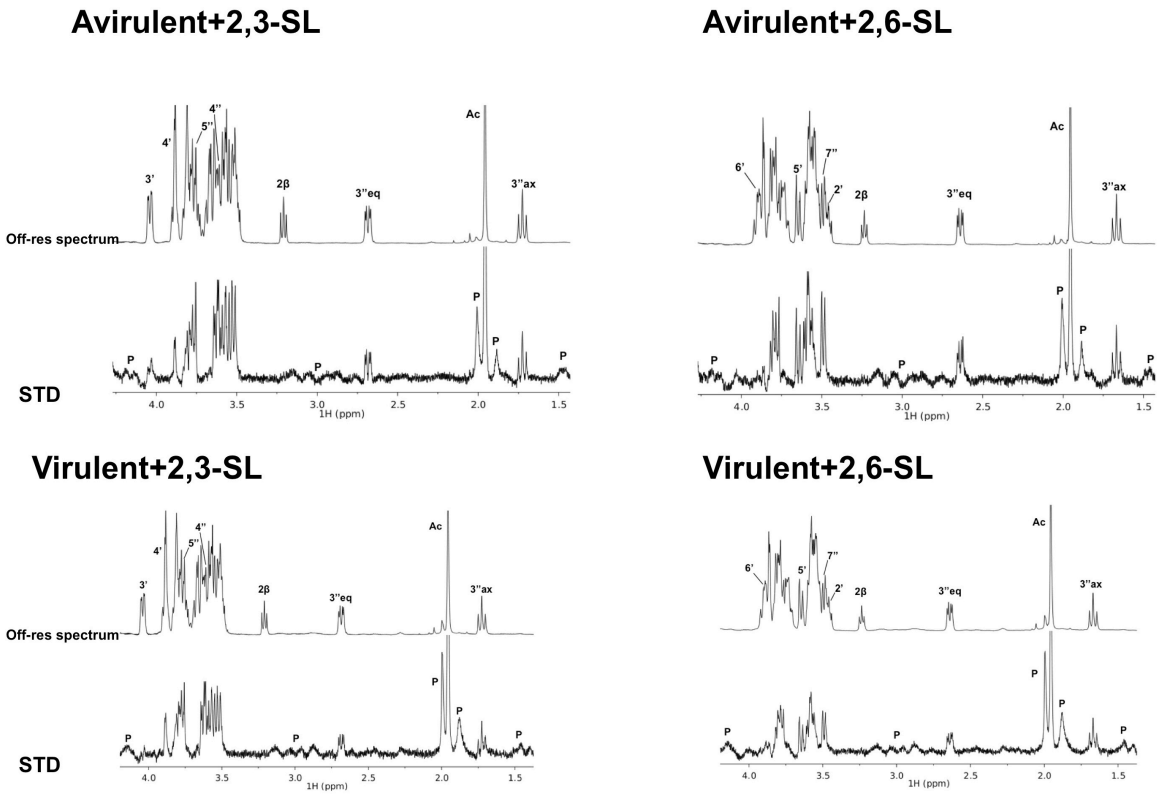


Figure 28: STD experiment performed on TAdV-3 fibre heads. In each section, upper spectrum indicates off-resonance (reference) spectrum, with labels indicating the assignments for a number of representative ligand signals. STD spectrum (up-scaled 100X) is shown on the lower half. Peaks marked with "P" correspond to signals belonging to the protein.

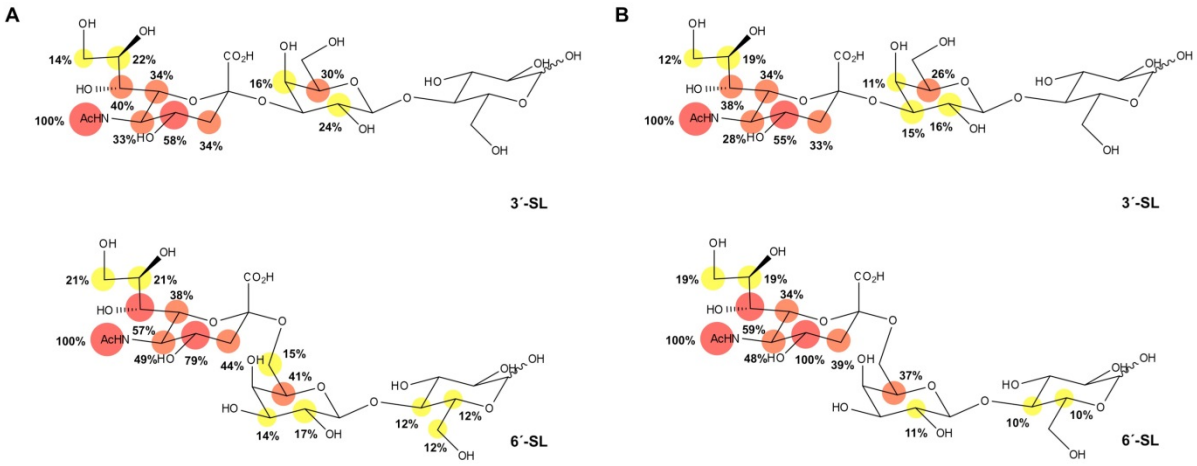


Figure 29: Epitope mapping of 2,3'- and 2,6'-sialyllactose. STD positive peak contribution by 2,3'- and 2,6'-sialyllactose in presence of (A) virulent TAdV-3 fibre head protein and (B) avirulent TAdV-3 fibre head protein with labels indicating the STD intensity for each signal, relative to the STD intensity for the 5''-N-acetyl peak. Red circles: I > 50%, orange circles: 50% I > 30%, yellow circles: 30% > I > 10%. Note the highest degree of saturation observed at the sialic acid part of the sugars in both cases.

STD-NMR was also performed with xylose, but no interaction with the proteins was observed. The chemical properties of slide surfaces in glycan microarrays as well as different linkers attaching glycans to their carrier molecules can affect the way a glycan is being presented and can impact target protein-ligand interaction (Kilcoyne et al., 2012). I speculate that this disagreement between glycan microarray profiling and STD-NMR results may be due the presentation of the carbohydrates and the linker between xylose and BSA.

To better understand the strength of these interactions, surface plasmon resonance (SPR) was used to determine the affinity of the TAdV-3 fibre head-sialyllactose interaction. Only one carbohydrate, 2,3'-sialyllactose, was selected for this study as it shares the same contact atoms with 2,6'-sialyllactose. The sugar was immobilized on a streptavidin sensor chip using its biotin tail, avirulent and virulent fibre head protein solutions were passed over the chip in different concentrations and protein-ligand association was monitored. Due to weak binding strength, the equilibrium state was not achieved for either protein. A subsequent increase in association time gave the same outcome, so it was not possible to calculate reliable kinetic parameters of the proteins. However, sensograms clearly showed specific interaction of both proteins with the carbohydrate (Figure 30). Thus, isothermal titration calorimetry (ITC) was employed to determine binding parameters of 2,3'-sialyllactose binding to avirulent and virulent variants of the TAdV-3 fibre head protein.

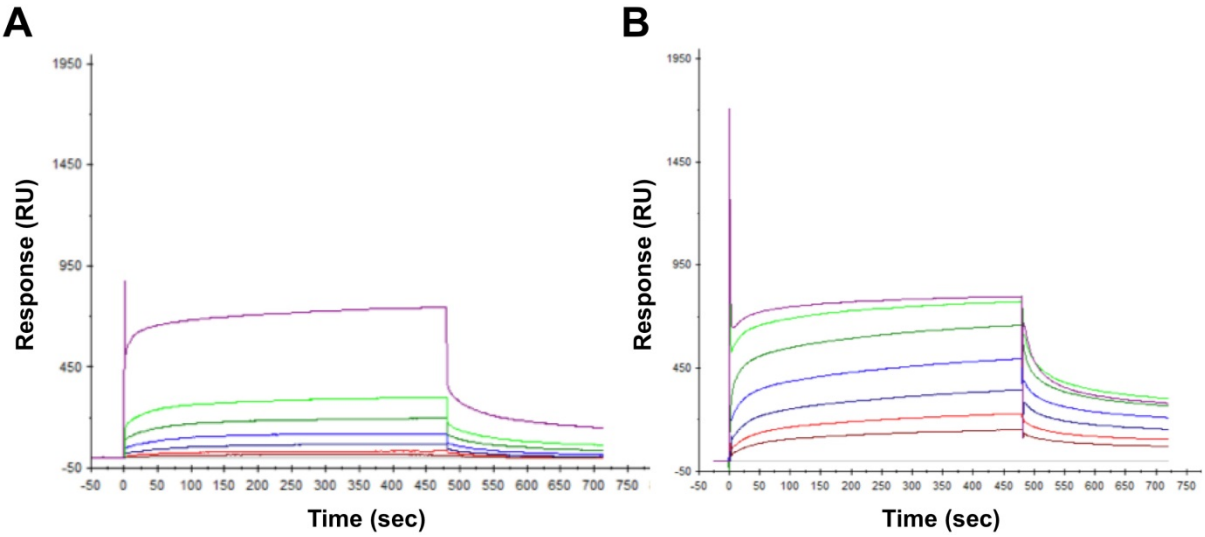


Figure 30: SPR analysis of 2,3'-sialyllactose binding to (A) avirulent and (B) virulent TAdV-3 fibre head proteins. Sensogram indicates protein concentration in µg/ml used during experiments: 7.5 (brown), 15 (red), 30 (deep blue), 60 (light blue), 120 (deep green), 240 (light green) and 480 (purple).

ITC results showed that both proteins bound to 2,3'-sialyllactose with mM affinity (kd value between 3.5-5.5 mM; Table 14; Figure 31). The variation of the heat released per mole of ligand injected with the sugar/protein ratio can be described with a model of one binding site per protein monomer subunit. The thermodynamic parameters derived from curve fitting suggested slightly different ways of interaction for the avirulent and virulent forms of the TAdV-3 fibre head protein, since both the enthalpic and the entropic contributions of the sugar binding differ (Table 14). The lower enthalpic stabilization of the complex in the virulent form is partially compensated by a more favorable entropic contribution.

Table 14: ITC-derived thermodynamic parameters for 2,3'siallylactose binding to TAdV-3 virulent and avirulent fibre head proteins. Measurements (MES bufffer, pH 6.0, 25ºC) were fitted assuming three independent sites per protein trimer.

Protein	K_b (M^{-1})	K_D (mM)	- ΔH (kcal/mol of sites)	- ΔS (cal/mol of sites/deg)
TAdV-3 (virulent)	182	5.4	1.83	- 4.2
TAdV-3 (avirulent)	275	3.6	4.43	3.70
Lys421Ala[a]	n.o.[b]	-	-	-
Lys439Ala[a]	311	3.2	3.9	1.5
Arg368Ala[a]	310	3.2	3.86	1.55
Glu392Ala[a]	~ 50	~20	n.d.	n.d.
Asn407Ala[a]	360	2.7	4.66	3.93

[a]TAdV-FH avirulent mutants. [b]Not observed. [c]Not determined.

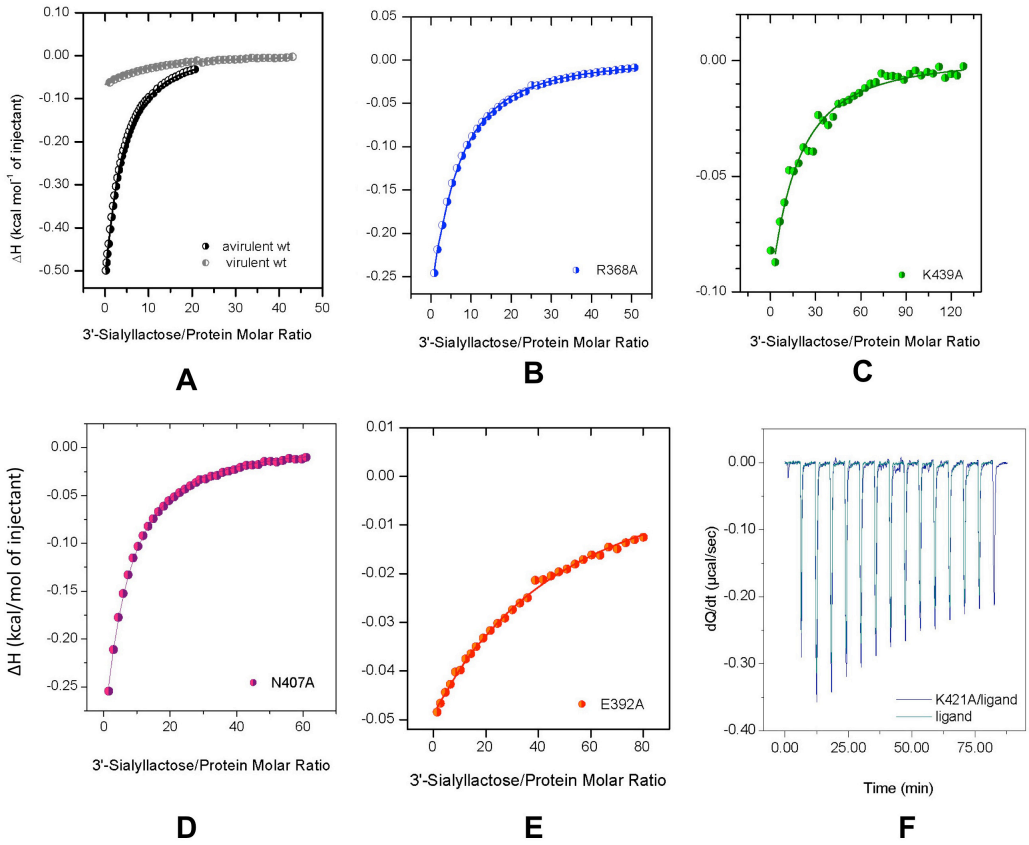


Figure 31: ITC profiles of the TAdV-3 fibre head when analysed with 2,3'-sialyllactose. Isotherm obtained from avirulent and virulent fibre heads (A) and that from single point mutants; (B) Arg368Ala, (C) Lys439Ala, (D) Asn407Ala and (E) Glu392Ala. (F) Titration curve for K421A mutant showing heat changes obtained when ligand alone and with protein was injected.

4.2.8 Co-crystal structures with 2,3'- and 2,6'- sialyllactose

Soaking of avirulent and virulent TAdV-3 fibre head crystals were performed with 2,3'- and 2,6'-sialyllactose as described in Material and Methods. The ligand bound structures were solved by molecular replacement using the native model. Data collection and refinement statistics are given in Table 15.

Table 15: Crystallographic data measurement and refinement statistics for ligand bound crystals

	2,3'-Sialyllactose		2,6'-Sialyllactose	
	Avirulent	Virulent	Avirulent	Virulent
Data collection				
Synchrotron and beamline	ESRF BM30	ESRF BM30	ESRF BM30	ESRF BM30
Detector	ADSC Q315r	ADSC Q315r	ADSC Q315r	ADSC Q315r
Wavelength (Å)	0.9797	0.9797	0.9795	0.9795
Crystal-detector distance (mm)	291.0	291.0	291.9	291.8
Oscillation range (°)	1.0	1.0	1.0	1.0
Number of images	67	150	80	51
Space group	*I*23	*I*23	*I*23	*I*23
Cell edge (Å, a=b=c)	98.36	98.82	98.45	98.86
Mosaicity (°)	0.6-0.7	1.1-1.2	0.5-0.7	0.5-0.7
Resolution range (Å)	49.1-2.5 (2.64-2.5)*	49.4-2.5 (2.64-2.50)*	49.2-2.2 (2.32-2.20)*	40.3-2.3 (2.42-2.30)*
Reflections	5623 (811)	5711 (815)	8227 (1177)	7109 (1037)
Multiplicity	7.7 (7.9)	16.6 (17.2)	9.3 (9.6)	5.9 (6.0)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	98.3 (99.9)
Mean <I/s(I)>	14.2 (2.4)	23.9 (4.2)	18.9 (4.1)	15.5 (2.7)
R_{sym} (%)†	10 (91.9)	7.9 (74.0)	6.9 (55.8)	6.4 (65.8)
V_M ($Å^3Da^{-1}$)‡	1.98	2.0	1.99	2.0
Wilson B ($Å^2$)	45.0	50.5	41.1	53.8
Refinement				
Resolution range (Å)	40.19-2.50 (2.6-2.5)	40.34-2.50 (2.6-2.5)	49.27-2.20 (2.3-2.2)	34.95-2.30 (2.4-2.3)
Reflections used	5355 (768)	5434 (384)	7841 (1118)	6781 (493)
Reflections used for R_{free}	261 (43)	263 (16)	383 (56)	328 (21)
R-factor	0.193 (0.267)	0.194 (0.223)	0.177 (0.225)	0.206 (0.318)
R-free	0.254 (0.346)	0.273 (0.368)	0.215 (0.32)	0.261 (0.400)
Protein/phosphate/ligand/water atoms	1092/5/21/30	1091/5/21/24	1092/5/21/52	1091/5/21/23
<B>overall/protein/ion/ligand[a] /water ($Å^2$)	51.2/51.12/87.7/76.9/ 41.3	57.1/57.6/44.6/60.0/4 9.25	47.1/46.7/41.0/80.6/ 45.6	64.8/63.9/129.4/139.0 /55.3
Ramachandran statistics (%)	90.4/98.5	97.1/100	97.1/100	95.6/97.8
r.m.s.d. bonds (Å)/angles (°)	0.010/1.5	0.014/1.5	0.007/1.2	0.009/1.4

* Values in parentheses are for the highest resolution bin (where applicable).

† $R_{symm} = \Sigma_h\Sigma_i|I_{hi}-\langle I_h\rangle|/\Sigma_h\Sigma_i|I_{hi}|$, where I_{hi} is the intensity of the *i*th measurement of the same reflection and $\langle I_h\rangle$ is the mean observed intensity for that reflection.

‡ According to Matthews (1968)

[a] Refined with occupancy of 1.0

The structures revealed two possible binding sites for the ligands on the protein. The ligand was found making contacts with two protein monomers, each of them belonging to independent asymmetric units. The pattern of interaction arosed two possibilities: the interaction mediated by one of the monomeric site is an artifact of crystal contact or both monomeric sites actually contribute in ligand binding (Figure 32A). Further experiments were conducted to establish the actual binding site. I identified key interacting residues of the protein and prepared single point mutants for each to test their effect on 2,3'-sialyllactose recognition by ITC. The substitution of residues Lys439, Arg368 and Asn407 by alanine (all part of site 1 as indicated in figure 32B) had very little impact on the binding thermodynamics, although a small increase in affinity was found for the three mutants accompanied by certain enthalpic/entropic compensation in the energetic of complex formation (Figure 31; Table 14). In contrast, the substitution of Glu392 with alanine drastically reduced the affinity of the mutant for 2,3'-sialyllactose and a similar replacement at position 421 (Lys421Ala mutant) rendered carbohydrate binding undetectable (Table 14; Figure 31). These two residues constitute binding site 2 (Figure 32B). These mutagenic effects on ligand binding were validated by STD-NMR and similar results were obtained (Figure 33).

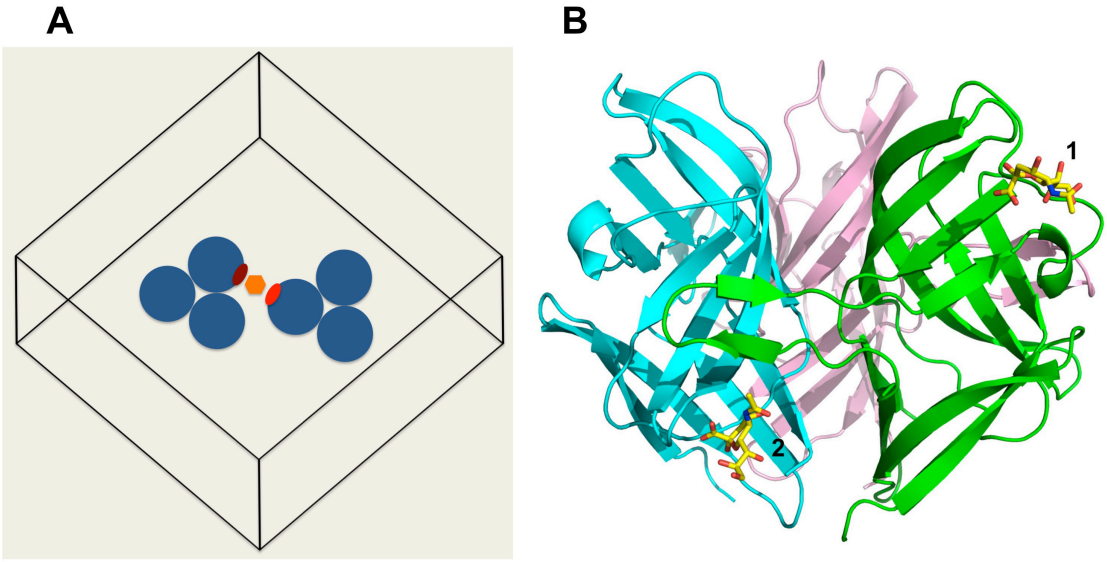


Figure 32: Interaction between 2,3'- or 2,6'- sialyllactose and TAdV-3 fibre head within crystals. (A) Schematic representation of the interactions. Ligand located between two monomers, each belonging to independent asymmetric units, is shown in orange while binding site over each monomer is indicated in brown and orange ovals. Biologically relevant trimers are generated by the crystallographic three-fold symmetry axis. (B) The binding sites mapped over the trimeric fibre head.

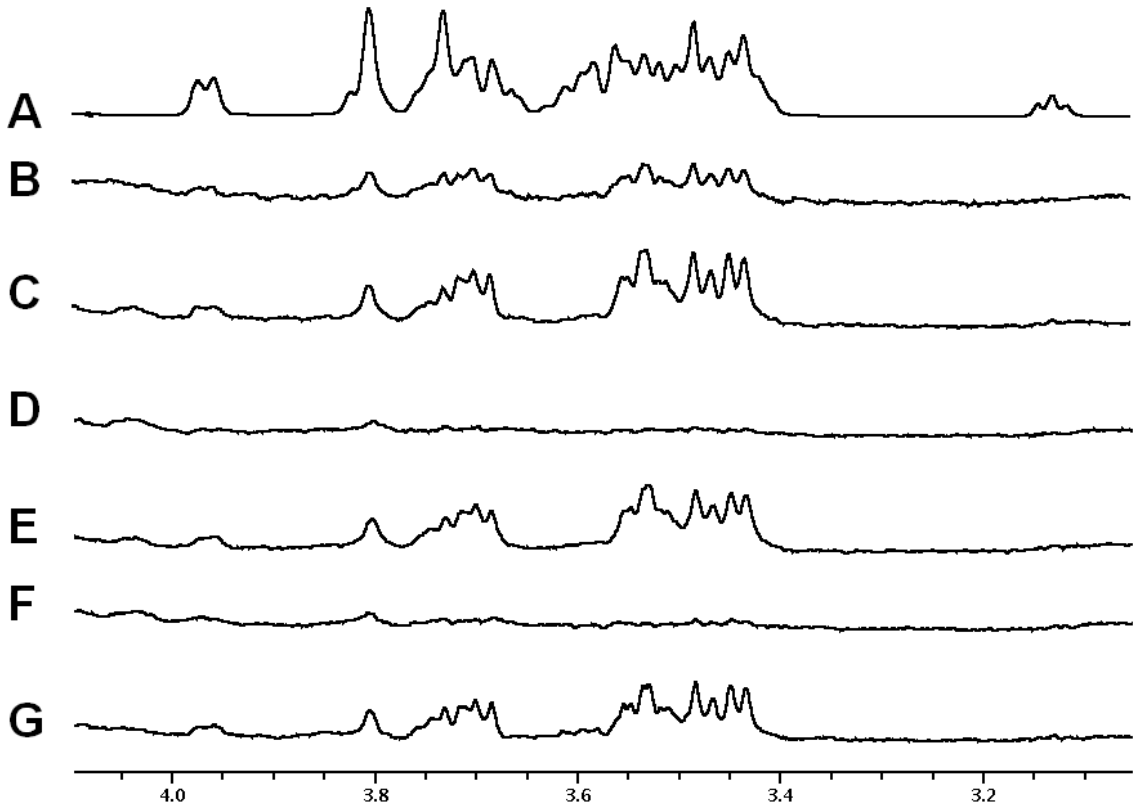


Figure 33: STD experiments performed on 2,3'-sialyllactose in the presence of different TadV-3 fibre head mutants. A: Off-resonance (reference) spectrum. B: STD spectrum with wt virulent protein. C: STD spectrum with R386A mutant protein. D: STD spectrum with E392A mutant protein. E: STD spectrum with N407A mutant protein. F: STD spectrum with K421A mutant protein. G: STD spectrum with K439A mutant protein.

Site 2 is located towards the lower half and close to the interface between two monomers of the trimeric assembly (Figure 32B). Each monomer of the trimer contributes a site for ligand binding and there was no evidence for cooperativity. In all cases, unbiased densities were observed only for the sialic acid units and hence lactose parts were not modelled (Figure 34). This was expected based on the findings of glycan array and STD-NMR, where the sialic acid part was shown to be more important for binding. The two protein variants have shown differential preference for the ligands where the densities for 2,3'-sialyllactose were slightly stronger comparing to that for 2,6'-sialyllactose and, for both ligands, avirulent fibre head has shown a higher binding affinity. Average temperature factors of the ligands were comparable to those of proteins in all but one case (Table 15). There were no significant differences between ligand-free and ligand-bound models and overall r.m.s.d. values of 0.3-0.4 Å was obtained upon superposing them.

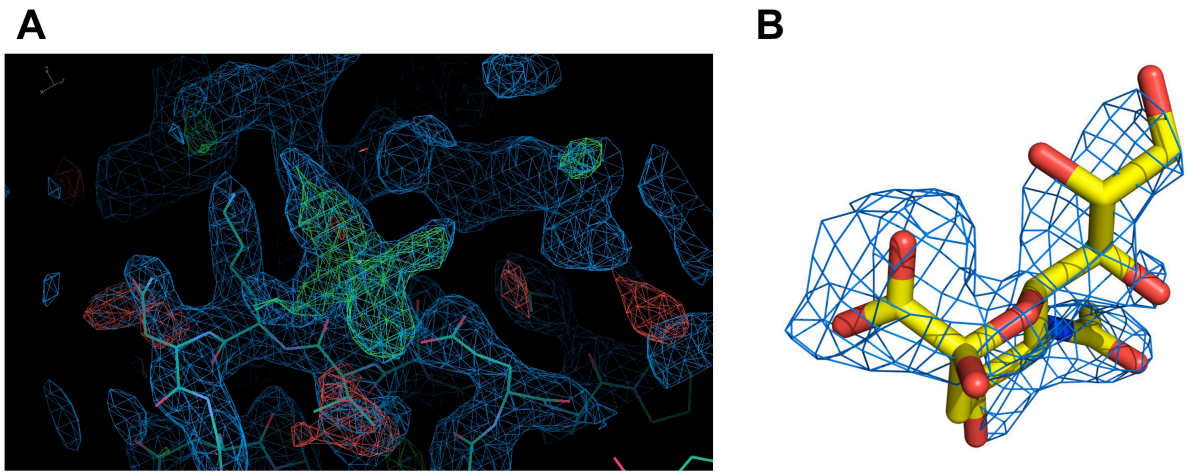


Figure 34: Experimental electron density obtained for 2,3'-sialyllactose sugar bound to avirulent TAdV-3 fibre head. (A) A section of unrefined experimental map showing density for sialic acid unit. 2Fo-Fc density (blue) is contoured at 1 σ while positive Fo-Fc density (green) is contoured at 3 σ. (B) Unbiased density observed for sialic acid. The 2Fo-Fc map covering the ligand is calculated from structure factors obtained prior to ligand fitting and contoured at 1 σ.

The structures revealed the sialic acid binding site of TAdV-3 fibre head (Figure 35A,C). Participating protein residues are same in both fibre head variants with Thr419, Val420, Lys421, Asn422 and Gly423, which are located on the H strand and HI loop of the monomer, constituting the binding site. The surface exhibits a strong electropositive charge distribution that drives the binding of negatively charged sialic acid molecule (Figure 35B). Ligands in their energy minimum chair conformation readily present axial carboxylic group and make several hydrogen and hydrophobic interactions with above mentioned residues. ε-amino group of Lys421 makes a close distance hydrogen bond with carboxylic group O8 of sialic acid while its α, δ and ε carbons are engaged to sialic acid via several hydrophobic interactions (Figure 35D). Thr419 and Val420 mainly contribute hydrophobic interactions while Asn422 and Gly423 makes hydrogen as well as hydrophobic bonds with the sugar molecule. Apart from these contacts, Glu392 located over the adjacent DG loop of the same monomer plays a crucial role in overall stabilization of the ligand by making two hydrophobic contacts with C10 and C11 of sialic acid. Jointly with Thr419 and Val420, Glu392 stabilizes the N-acetyl group of sialic acid (Figure 35D). Both Lys421 and Glu392 engage terminal regions of sialic acid and thus provide the molecule an end-to-end attractive force. As previously stated, mutation of either of Lys421 or Glu392 heavily disrupts binding of the ligand exemplifying their role in establishment of the interaction.

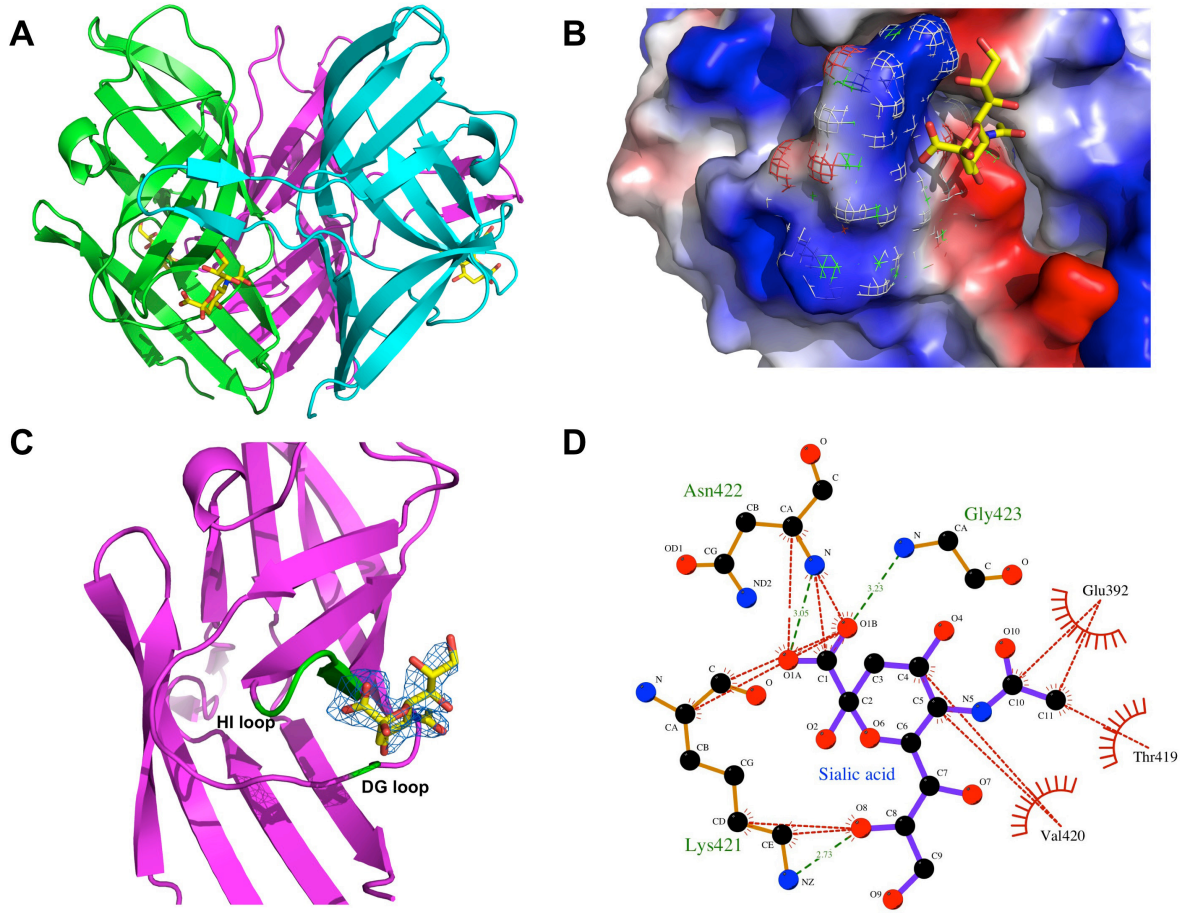


Figure 35: Structure of the avirulent TAdV-3 fibre head bound to 2,3'-sialyllactose. (A) Side view and of trimeric fibre head bound to the ligand. Trimer was generated by crystallographic 3-fold symmetry axis. (B) Surface representation of the ligand bound structure. Binding site surface is depicted in mesh. Notice the orientation of carboxylic and N-acteyl groups of sialic acid (C) Close view of the ligand binding site. The bound sialic acid unit is displayed with its 2Fo-Fc electron density contoured at 1 σ and calculated as in figure 34B. Interacting regions of HI and DG loops are shown in green color. (D) Fibre head-2,6'-sialyllactose-interaction plot. The central sialic acid unit is labeled. Residues forming hydrogen bonds are indicated in green while those making hydrophobic contacts are in black. Dashed green and red lines denote hydrogen and hydrophobic bonds, respectively.

4.2.9 TAdV-3 fibre head has a sialyllactose-binding site different from other adenoviruses

Both variants of the TAdV-3 fibre head can bind sialylated glycans as shown by solution experiments and ligand bound structures. However, it's intriguing to observe that one protein (aviruelnt) engages the ligand with a slightly higher affinity than the other (virulent) especially when both variants are almost identical in structure. The only observable difference in their structures lies in the beta-hairpin arm region, which shows movement and is slightly bended upwards in case of virulent. This movement might not be very influential

upon the overall trimeric structure but can certainly bring local conformational changes. The ligand-binding site of the protein is situated very close to the beta-hairpin region and I speculate that this might be the reason why there is variation in ligand attachment. Other possibility could be the requirement of a more complex glycan for attachment of virulent fibre. It was observed in case of HAdV-37 that the binding of its fibre head with cellular glycosylated receptor (with two terminal sialic acid moieties) was many folds stronger in comparison to simple sialyllactose (Nilsson et al., 2011). So, if indeed the virulent fibre head binds a glycan cellular receptor under physiological conditions, then perhaps it requires a more complex and branched molecule in order to make binding feasible and effective.

Many viruses can bind sialic acid via their receptor-recognising proteins. These includes VP1 and VP4 proteins of polyomavirus and rotavirus respectively, influenza virus haemagglutinin, and fibre proteins of adenovirus and reovirus besides many others (Reiter et al., 2011; Seiradake et al., 2009; Burmeister et al., 2004; Dormitzer et al., 2002b; Stehle and Harrison, 1997; Weis et al., 1988). Although the nature of interaction remains charge-dependent, the sialic acid binding site over these proteins varies significantly in term of location. Among adenoviruses itself, different fibre heads have varying silaic acid binding locations. Figure 36 shows a comparison between sialic acid binding locations of TAdV-3, CAdV-2 and HAdV-37 fibre head domains. CAdV-2 and HAdV-37 have these sites located over top of their trimeric assemblies while sialic acid attaches close to the monomeric interface and towards the bottom of the TAdV-3 fibre head trimer (Figure 36). Mammalian reovirus fibre shares a close resemblance to TAdV-3 fibre head with conserved topological features. Its receptor binding protein sigma1 can also bind sialic acid, however, in this case the viral protein uses its shaft domain to engage the molecule (Reiter et al., 2011).

Apart from differences in their binding locale, some differences can also be observed in nature of bonds involved in these viral protein-sialic acid interactions, which mainly occurs due to type of participating amino acid residues. For instance, HAdV-37 and VP1 make salt bridge connections with the incoming ligands whereas salt bridges are not present in TAdV-3 fibre head, influenza virus haemagglutinin and VP4 interactions with sialic acid. Main forces involved in latter cases are hydrogen bonds and hydrophobic interactions. However, conservation of sialic acid epitopes can easily be noted. Carboxyl group of sialic acid is readily presented for binding in case of VP1, VP4 and HAdV-37, a feature also observed in case of TAdV-3 fibre head-sialic acid interaction.

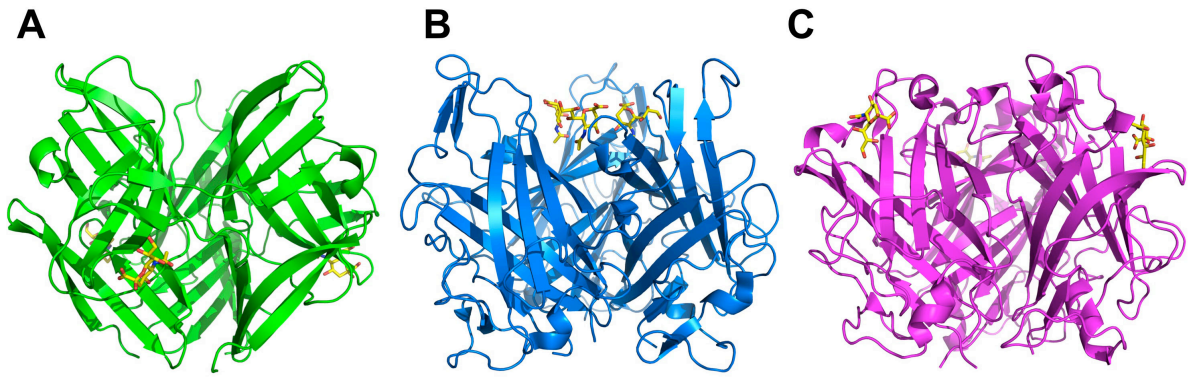


Figure 36: A comparative view of sialyllactose binding adenovirus fibre heads and their binding sites. (A) TAdV-3 fibre head bound to sialyllactose acid molecule. The site is located towards the lower half and close to the monomeric interface of the trimer. (B) HAdV-37 fibre binds sialic acid over top of its assembly, very close to the 3-fold axis. (C) Similar to HAdV-37 fibre head, sialic acid binding site of CAdV-2 is also located on the top of the trimeric assembly, however it's away from the 3-fold symmetry axis. In all depictions density is shown only for the sialic acid unit of sialyllactose.

4.2.10 Conclusion

The atomic structures of the fibre head of the avirulent and virulent variants of TAdV-3 are described. This is the first Siadenovirus for which the structure of the fibre head has been determined. The structure revealed a strongly positively charged surface, which may be important for host cell interaction. Interactions of fibre head variants with 2,3'-sialyllactose and 2,6'-sialyllactose have been shown by several biophysical methods and co-crystal structures were determined. The co-crystal structures revealed the sialic acid binding site of the TAdV-3 fibre head, which is located close to monomer-monomer interface and bears a postive surface charge distribution. A beta-hairpin insertion in the C-strand contacting a neighbouring monomer is unique to the TAdV-3 fibre head structure. The lack of sequence similarity to other Siadenovirus fibre heads precludes me from suggesting that other siadenovirus fibres also contain this beta-hairpin, although this is certainly possible. The fibre head shares its ABCJGHID beta-barrel topology with reovirus fibre heads, while certain bacteriophage baseplate receptor-binding proteins and other adenovirus have a beta-sandwich with separate ABCJ- and GHID-sheets. As reoviruses are double-stranded RNA viruses and there seems to be no possible direct evolutionary relationships between adenoviruses and reoviruses, it is unclear if (and how) the fibre genes may have passed between them or whether the evolutionary process involved has been convergent rather than divergent. Knowledge of the structure may also allow the design of chimeric adenoviruses incorporating the TAdV-3 fibre head and investigating their tropism, which in turn may lead to vaccination or gene therapy vectors that may target specific cell types.

4.3 Crystal structure of lizard adenovirus 2 short fibre head domain

Lizard adenovirus type 2 (LAdV-2) is the first characterized Atadenovirus that has two distinct fibres on its virion (Penzes et al., 2014). The fibre-1 gene of LAdV-2 codes for a polypeptide of 331 amino acids with a monomeric weight of 34.7 kDa whereas the predicted length for the fibre-2 gene product is 433 amino acids with a molecular weight of 45.7 kDa. The two proteins differ significantly with each other and have a sequence identity of just 20%. I determined the high-resolution crystal structure of the 101 residues long LAdV-2 short fibre head domain by molecular replacement technique using SnAdV-1 fibre head as search model, with which it shares 54% sequence identity. Crystals belonged to the *P*1 space group and diffracted synchrotron radiation to maximum resolution of 2.0 Å. The structure is very similar to the SnAdV-1 fibre head, albeit 6 residues shorter.

4.3.1 Purification and crystallization

4.3.1.1 Construction of expression vectors

The amino acid sequence analysis of the LAdV-2 short and long fibres showed characteristics of a typical adenovirus fibre protein with presence of beta-spiral repeats in central shaft domain (van Raaij et al., 1999b). The tail and shaft regions of short fibre (fibre-1) run till residue 210 and after a short linker the putative head domain could be presumed to start. The long fibre or fibre-2 is 102 residues longer than the short fibre and in this case the beta-spiral region is predicted to encompass residues 46-260. The remaining C-terminal portion of the protein will probably form the head domain. Based on these features, cloning was performed and expression vectors were generated for several N-terminal truncated versions of the fibres, including C-terminal fibre heads.

4.3.1.2 Protein purification and crystallization

Expression vectors for short and long fibre protein constructs were used to produce milligram amounts of soluble protein, which was purified by nickel-affinity chromatography followed by strong anion-exchange chromatography. In case of the short fibre, the putative head construct starting at residue 220 was highly soluble and the protein yield was approximately 25 mg from 1 litre of bacterial culture. On the other hand, the protein yields for the two long fibre constructs starting at residues 245 and 305 were relatively lower at around 15 mg/litre of bacterial culture.

Proteins were eluted from the nickel column between imidazole concentrations of 0.1-0.5 M. In anion-exchange chromatography, the construct 220 of short fibre eluted in two peaks around 0.2-0.5 M sodium chloride when a linear gradient was used, each of which contained pure or nearly pure fibre head protein (Figure 37). Interestingly, this behaviour is very similar to that of SnAdV-1 fibre head protein, which also elutes in two prominent peaks (see section 4.1.1.2). The peak fractions were pooled and kept separate for crystallization purpose. The putative head domain construct 245 of long fibre showed a single peak of pure protein during anion-exchange chromatography (Figure 38).

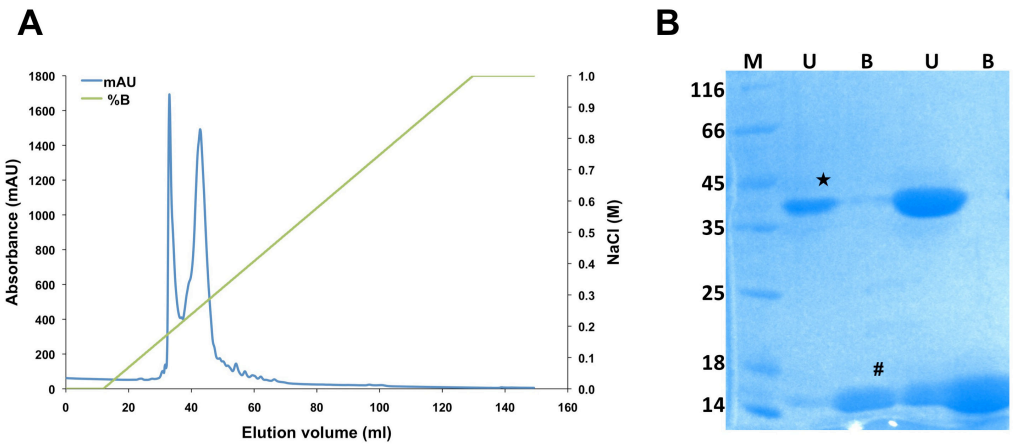


Figure 37: Purification of LAdV-2 short fibre head domain. (A) Protein elutes in two prominent peaks in anion-exchange (Resource Q) chromatography. In green the sodium chloride concentration gradient is indicated. (B) Separately pooled peaks are run over a SDS-gel. Lanes are marked M (molecular weight markers, with the size indicated in Kd), U (unheated sample) and B (sample heated to 95 ºC for at least 5 min). Trimeric protein in unheated fraction is shown by an asterisk symbol while hash indicates monomeric protein in the respective heated fraction.

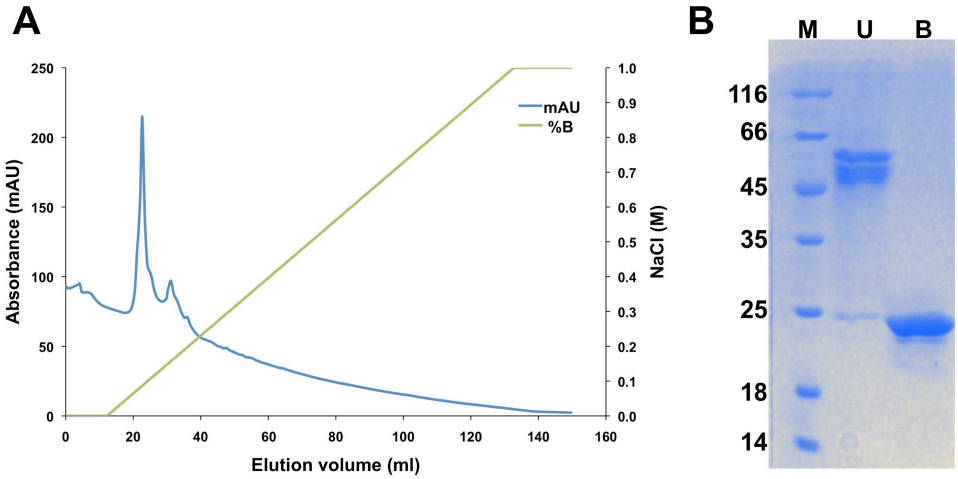


Figure 38: Purification of the LAdV-2 long fibre construct 245. (A) Chromatogram showing elution pattern during anion-exchange chromatography. In green the sodium chloride concentration gradient is indicated. (B) SDS-PAGE run of peak fraction pooled after anion-exchange chromatography. Lanes are marked M (molecular weight markers, with the size indicated in Kd), U (unheated sample) and B (sample heated to 95 ºC K for at least 5 min).

Crystals of short fibre head were obtained in four of the conditions, two of which resulted in good quality data sets. Condition 1 contains 20% (w/v) PEG 3350 and 0.2 M tri-potassium citrate, while condition 2 contains 0.1 M Tris-HCl pH 8.0 and 30% (v/v) PEG 600. Plate like crystals appeared in condition 1 and belonged to space group *P*1. They diffracted to a maximum resolution of 2 Å (Figure 39A; Table 16). They are expected to contain four trimers in the asymmetric unit, leading to a solvent content of 26%. Thick needle-shaped crystals appeared in a condition containing 0.2 M tri-Potassium Citrate, 20% w/v PEG 3350, which belonged to space group $P3_221$ and diffracted to 2.8 Å (Figure 39B). In this case, there was 1 trimer in the asymmetric unit with solvent content of 31.8 %.

Crystals of long fibre construct 245 took around 15 days to appear and could only be obtained in a single condition consisting of 0.2 M di-ammonium tartrate, 20% w/v PEG 3500. Diffraction quality of these crystals was very poor and obtained data was not useful.

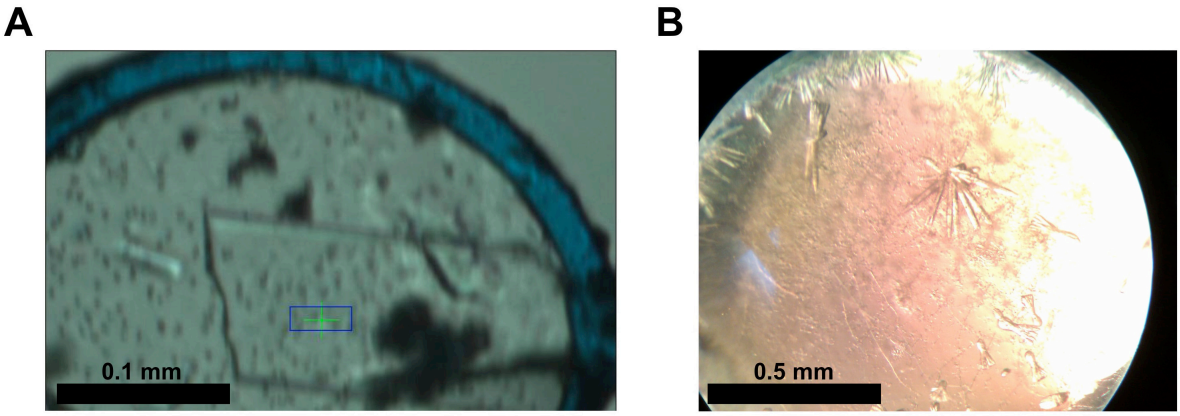


Figure 39: Crystals of lizard fibre heads. (A) Lizard short fibre head crystals belonging to *P*1 space group. (B) Thick needle shaped crystals of short fibre head domain in $P3_221$ space group.

Table 16: Crystallographic data measurement and refinement statistics for LAdV-2 short fibre head crystals

	Crystal form 1
Data collection	
Constructs (residues)	220-331
Synchrotron and beamline	SOLEIL, Proxima 2A
Detector	ADSC Q315r
Wavelength (Å)	0.9801
Crystal-detector distance (mm)	361
Oscillation range (°)	0.5
Number of images	400
Space group	*P*1
Cell edge (Å, a, b, c)	54, 79.5, 83.3
Cell angles (α, β, γ)	116.76, 90.10, 92.23
Mosaicity (°)	0.7-0.8
Resolution range (Å)	50.00-2.00 (2.11-2.00)*
Reflections	78829 (11467)
Multiplicity	2.1 (2.1)
Completeness (%)	95.3 (94.9)
Mean <I/s(I)>	5.5 (2.3)
R_{sym} (%)†	10.5 (35.0)
V_M ($Å^3Da^{-1}$)‡	1.66
CC Imean	0.984 (0.835)
CC anom	-0.431 (N.D)
Wilson B ($Å^2$)	13.8
Refinement	
Resolution range (Å)	19.86-2.00 (2.1-2.0)
Reflections used	77329 (5837)
Reflections used for R_{free}	3145 (453)
R-factor	0.221 (0.282)
R-free	0.279 (0.257)
Protein/water atoms	10535/943
Overall B ($Å^2$)	21.35
Ramachandran statistics (%)	97.1/100
r.m.s.d. bonds (Å)/angles (°)	0.009/1.4

* Values in parentheses are for the highest resolution bin (where applicable).
†$R_{symm} = \Sigma_h\Sigma_i|I_{hi}-\langle I_h\rangle|/\Sigma_h\Sigma_i|I_{hi}|$, where I_{hi} is the intensity of the *i*th measurement of the same reflection and $\langle I_h\rangle$ is the mean observed intensity for that reflection.
‡ According to Matthews (1968)
N.D. = Not determined

4.3.2 Structure determination of the LAdV-2 short fibre head domain

The LAdV-2 short fibre head domain has a close sequence relationship with the SnAdV-1 fibre head: 54% identity. The data set obtained from *P*1 space group crystal was used to determine the structure of LAdV-2 short fibre head. Molecular replacement was performed with a trimeric model of the SnAdV-1 fibre head. Four copies of the search ensemble were placed in the asymmetric unit. Automatic tracing and manual completion resulted in a model consisting of 103 amino acids in each chain (228-331) and 943 water molecules. Amino acids from the N-terminal purification tag could not be modelled in the structure due to lack of density, as was also the case for residues 220-227 of each of the chains. The final model was refined using NCS-restraints to 2 Å with good geometry and Ramachandran statistics (Table 16). The head domain appears to start at residue 230, while residues 220-229 are part of the linker between shaft and the head domain.

4.3.3 General features of the LAdV-2 short fibre head

The LAdV-2 short fibre head is the smallest adenovirus fibre head known so far, with only 101 residues per monomer. The trimeric assembly is 3.3 nm tall with a cross-sectional diameter of around 4.5 nm and each monomer has a surface area of 6.04 x 10^3 $Å^2$ out of which 23% gets buried upon trimer formation. Given the 54% amino acid sequence identity between the head domains of SnAdV-1 fibre and LAdV-2 short fibre, we were expecting to see a good degree of structural similarity. As expected, the obtained structure is virtually identical and resembles the SnAdV-1 fibre head in every aspect (Figure 40B). Both fibre heads are much smaller than other known adenovirus fibre head structures. The surface charge distribution of LAdV-2 short fibre head is of mixed nature with presence of both postive and negative patches of intermediate strength, a feature also observed in SnAdV-1 (Figure 40A).

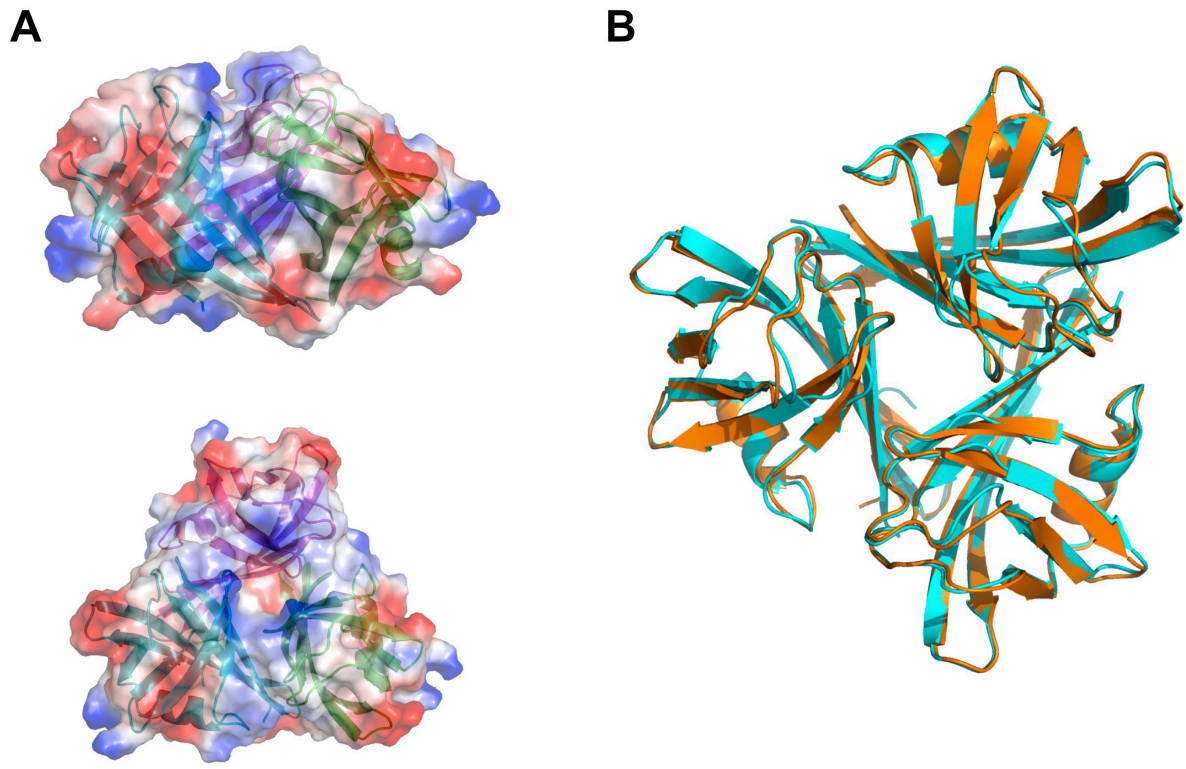


Figure 40: Structure of the LAdV-2 short fibre head. (A) Qualitative electrostatic surface diagram of the LAdV-2 short fibre head. (A) Side (upper section) and top (lower section) views of the trimeric assembly. (B) Structural superposition onto SnAdV-1 fibre head domain (orange).

4.3.4 Conclusion

The crystallographic structure of LAdV-2 short fibre head is determined and refined to a maximum resolution of 2Å. This is the second Atadenovirus protein for which atomic structure has been determined, the first being SnAdV-1 fibre head domain described in part 4. The structure is very similar to that of the SnAdV-1 fibre head, which was expected due to the 54% identity in sequence. Attempts were also made to crystallize the long fibre head, which only has 11.6% sequence identity to the short fibre head, but well-diffracting crystals could not be obtained. Future studies will be aimed at obtaining information on the receptor-binding properties of these proteins.

4.4 Crystal structure of the carboxy-terminal fibre head domain of murine adenovirus 2

Murine adenovirus 2 (MAdV-2) is a member of the *Mastadenovirus* genus of adenoviruses, along many other mammalian adenoviruses, including all human adenoviruses. It is the second murine adenovirus for which genomic information is available. A putative fibre head construct and a longer construct with potential shaft domain repeats were expressed and purified. Well-diffracting crystals for the putative head domain were obtained and its structure was determined using single anomalous dispersion (SAD) and refined to 1.8 Å. The structure revealed two shaft domain repeats besides a complete head domain. The structure has conserved topological features comparable to other adenovirus fibre heads and its closest homolog is canine adenovirus 2 fibre head. Binding assays showed that MAdV-2 fibre head binds N-acetyl glucosamine (GlcNAc) and that contacts involve the acetyl group of the glycan. This suggests MAdV-2 may use GlcNAc-decorated molecules for its cellular attachment. No binding with the human CAR D1 domain could be observed.

4.4.1 Purification and crystallization

4.4.1.1 Construction of expression vectors

Secondary structure prediction of the 787 amino acid long MAdV-2 fibre showed presence of a N-terminal penton base binding sequence, a central beta-spiral repeats containing shaft domain and a virus distal C-terminal fibre head domain, although the sequence identity of the C-terminal part with other known adenovirus fibre heads was restricted to between 10-20% only. At sequence level the closest homologue for which structure is known was found to be human adenovirus 41 short fibre head with 16% identity. The genomic DNA of MAdV-2 fibre was subjected to two separate PCR reactions with the aim of amplifying C-terminal fibre regions between residues 517-787 and 586-787, in which the former construct was expected to contain part of the shaft and the complete head domain, while the latter construct would contain the head domain only. The amplified PCR products were cloned into the pET28a(+) vector, which provides the protein with N-terminal 6xHis and T7 tags. The constructs were sequenced and confirmed to be as expected.

4.4.1.2 Protein purification and crystallization

Following expression of the protein constructs fib(517-787) and fib(586-787) in *E. coli*, the proteins were subjected to a two-step purification protocol, where they were first purified by Ni-NTA metal affinity chromatography and later by strong anion exchange chromatography. Both C-terminal MAdV-2 fibre constructs showed strong binding tendency towards nickel, as evident by their chromatography profiles where they eluted at relatively high imidazole concentrations (Figure 41).

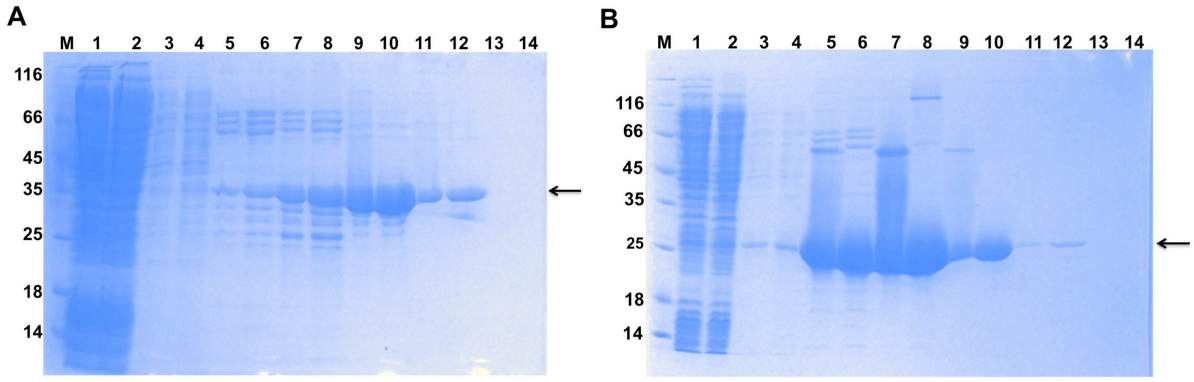


Figure 41: Metal affinity purification profiles for MAdV-2 fibre construct containing resdiues 517-787 (A) and residues 586-787 (B). Lane M indicates molecular weight marker; lanes 1 & 2: flowthrough unheated and heated respectively. Imidazole step gradient elution; lanes 3 & 4: 50 mM imidazole unheated and heated respectively, 5 & 6: 100 mM unheated and heated respectively, 7 & 8: 200 mM unheated and heated respectively, 9 &10: 300 mM unheated and heated respectively, 11 & 12: 400 mM unheated and heated respectively and 13 & 14: 400 mM unheated and heated respectively. Smaples were heated at 95 ºC for 5 min. Arrows indicate monomeric units of the trimeric proteins.

Although heating of fibre protein at high temperature is generally required in order to visualize its constituting monomers (Mitraki et al., 2006), partial dissociation of trimeric proteins takes place on SDS-PAGE in this case, as monomeric species were present in the unheated fractions (Figure 41). In anion exchange chromatography, proteins eluted as single peaks at around 0.1-0.4 M sodium chloride when a linear gradient of sodium chloride was used (Figure 42). From 1 litre of bacterial culture, the total protein yield was around 12 mg in both cases.

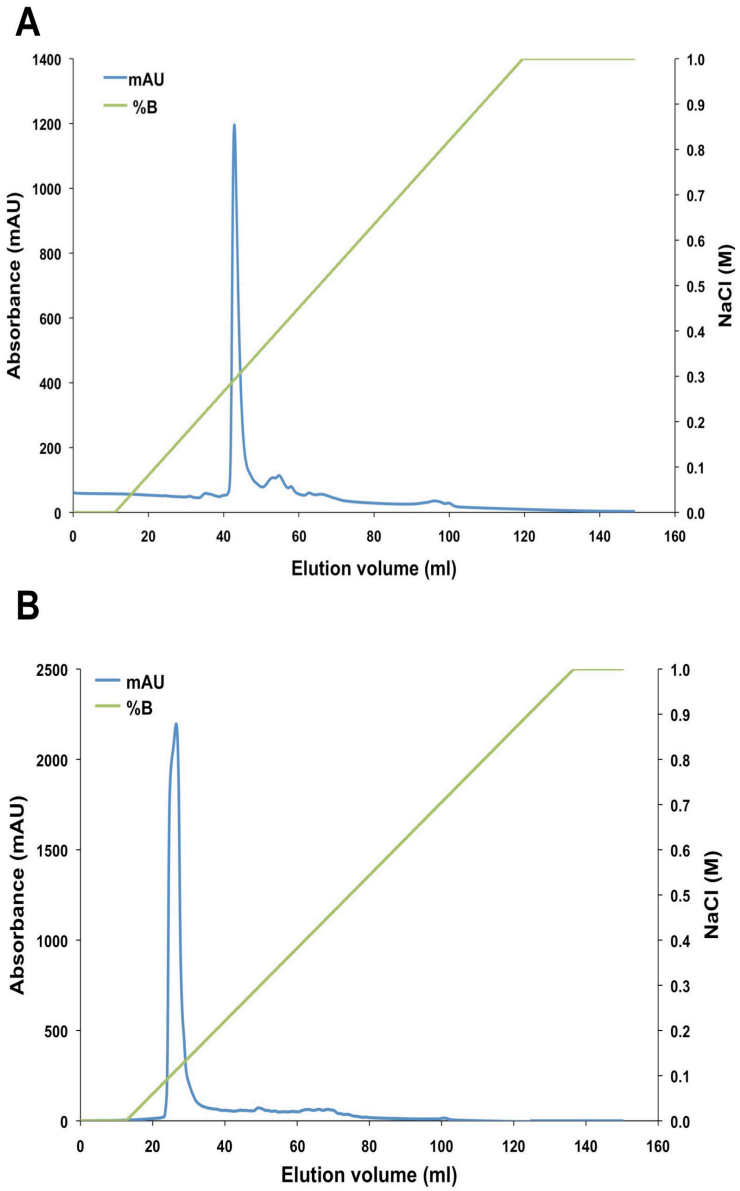


Figure 42: Anion-exchange chromatography profiles of MAdV-2 fibre protein constructs (A) fib(517-787) and (B) fib(586-787). Both fragments elutes in sharp, single peak at 0.1-0.4 M sodium chloride concentration. Green line indicates sodium chloride concentration gradient.

The longer fragment (containing residues 517-787) failed to produce crystals despite of extensive trials. This was perhaps due to presence of a part of a more flexible N-terminal domain, hindering crystallization. On the contrary, the crystals of the putative head domain (amino acids 586-787) were obtained in several conditions after 10-30 days. Plate-shaped crystals were obtained in when drops were setup in condition containing 4.0 M ammonium formate and 0.1 M HEPES sodium salt pH 7.5 (Figure 43A). They were found to belong to the orthorhombic space group $P2_12_12_1$. Heavy atom derivatization of these crystals was performed by soaking them with methyl mercury chloride. The soaked crystals tested for diffraction were in space group $C222_1$. A second crystal form appeared in monoclinic space

group $P2_1$ with a similar plate like morphology. In this case the reservoir solution contained 0.1 M magnesium chloride, 0.1 M HEPES-NaOH pH 7.5 and 10 % w/v PEG 4000 and 0.15 M ammonium sulphate, 0.1 M HEPES-NaOH pH 7.0, 20 % w/v PEG 4000 (Figure 43B). See Table 17 for crystal characteristics.

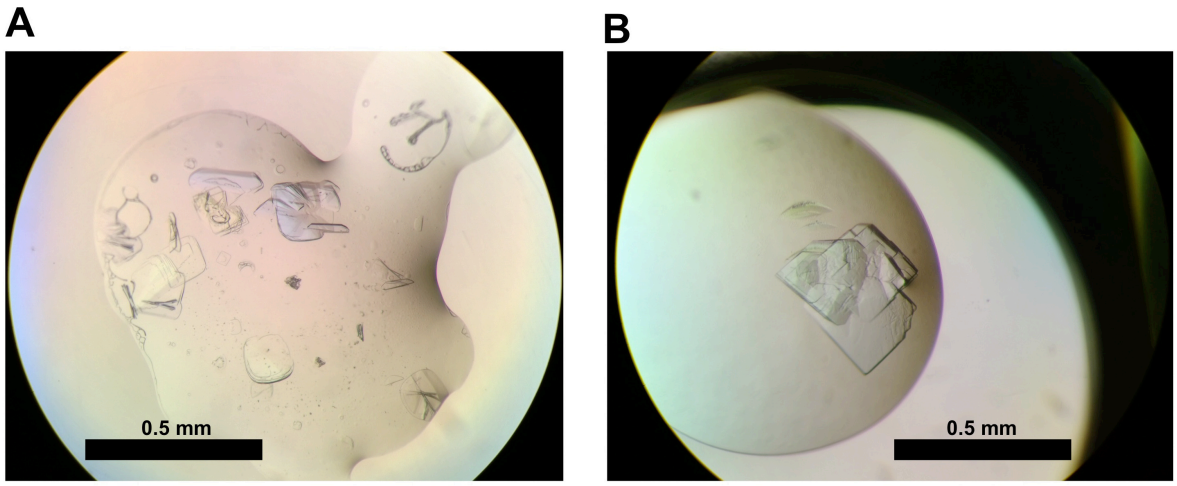


Figure 43: Crystals of MAdV-2 putative fibre head construct 586-787. (A) Plate like native crystals appeared in $P2_12_12_1$ or $C222_1$ space group. (B) The protein construct also crystallized in $P2_1$ space group with similar but more robust appearance. Bar measures approximately 0.5 mm.

4.4.2 Structure determination of the MAdV-2 fibre construct containing residues 586-787

The crystallized MAdV-2 fibre fragment has 16% sequence identity to its most homologous known structure as mentioned before. At such low sequence identity molecular replacement is unlikely to work (McCoy, 2007), therefore heavy atom derivatisation with methyl mercury chloride was performed. A single, complete and redundant data set was collected at a wavelength to maximize the anomalous signal from the incorporated mercury atoms. The crystal diffracted to a maximum resolution of 2.76 Å and the overall phasing power was above one up to 4.1 Å resolution. The asymmetric unit was expected to contain one protein trimer and therefore twelve mercury sites were expected in the asymmetric unit (each protein monomer contains four cysteine residues). All of these sites were identified successfully and refinement of these sites resulted in high-quality phases (Table 17).

Table 17: Crystallographic data measurement and refinement statistics for MAdV-2 fibre crystals

	Hg-derivative	Native 1	Native 2
Data collection			
Synchrotron and beamline	ALBA BL13-XALOC	ALBA BL13-XALOC	ESRF ID29
Wavelength (Å)	1.0024	1.0022	0.9762
Crystal-detector distance (mm)	521.5	327.0	551.7
Oscillation range (°)	0.2	0.2	0.1
Number of images	1440	648	1160
Space group	$C222_1$	$P2_1$	$P2_12_12_1$
Cell edge (Å, a, b, c)	92.11, 164.28, 96.43	93.89, 100.61, 95.46	55.87, 91.11, 128.92
Mosaicity (°)	0.1-0.2	0.4-0.6	0.2-0.3
Resolution range (Å)	96.4-2.7 (3.02-2.76)	29.6-1.8 (1.9-1.8)	34.0-1.7 (1.74-1.70)
Reflections	19199 (4473)	144440 (7093)	68912 (3722)
Multiplicity	11.7 (8.0)	2.8 (2.9)	3.5 (1.9)
Completeness (%)	99.7 (98.8)	98.2 (97.8)	94.3 (69.6)
Mean <I/s(I)>	19.6 (3.2)	6.8 (2.3)	6.3 (2.1)
R_{sym} (%)†	9.7 (55.2)	7.9 (33.4)	11.6 (23.3)
V_M ($Å^3Da^{-1}$)	2.35	2.60	2.11
Wilson B ($Å^2$)	51.5	16.8	14.5
Phasing			
Heavy atom sites	12 Hg		
Correlation coefficient (all/weak)	52.40 / 22.50		
Patterson figure of merit	14.24		
Correlation coefficient (E)	0.462		
R-cullis (anomalous, acentric)	0.887		
Phasing power (anomalous differences)	0.753		
FOM = cos (phase error) (acentric/centric)	0.2405 / 0.0758		
Solvent flattening (44.6% solvent content)			
R-factor (before/after)	0.47 / 0.45		
Overall corr. on $\|E\|^2$ (before/after)	0.5262 / 0.8837		
Correlation on $\|E\|^2$ / contrast (original/inverted)	0.2470 / 0.1310		
Refinement			
Resolution range (Å)	82.14-2.76 (2.8-2.7)	29.66-1.80 (1.9-1.8)	34.0-1.7 (1.8-1.7)
Reflections used	18197 (1269)	140517 (20295)	65280 (9409)
Reflections used for R_{free}	985 (69)	3644 (499)	3559 (540)
R-factor	0.207 (0.339)	0.205 (0.259)	0.191(0.266)
R-free	0.284 (0.400)	0.216 (0.269)	0.222 (0.273)
Protein/water atoms/glycerol	4491 / 32	9024 / 758 / 24	4577 / 579
<B> overall/protein/water/glycerol ($Å^2$)	57.45 / 58.98 / 40.93	25.41/25.23/31.48/31.76	19.87 / 19.80 / 27.87
Ramachandran statistics (%)	90.2 / 95.1	97.7 / 100	95.8 / 99.3
r.m.s.d. bonds (Å)/angles (°)	0.008 / 1.3	0.008 / 1.3	0.009 / 1.3

Values in parentheses are for the highest resolution bin (where applicable).

Solvent flattening improved the phases further and a clearly interpretable electron density map was obtained. Some discontinuity was observed when the chains were automatically traced as well as some instances of wrong amino acid assignments were noticed, which were corrected manually. The partially refined model consisted of 193 amino acid residues (594-787) from each chain. The remaining eight N-terminal residues of the protein as well the vector-supplied tags could not be modelled. This model was then used in molecular replacement search with the high-resolution native data set obtained for the $P2_12_12_1$ crystal form. Automatic plus manual model building resulted in a structure with 193 residues for each of the three chains and a total of 594 water molecules. Manual inspection of the model did not show any trace for additional N-terminal residues. The other native data set from $P2_1$ space group crystal was processed in a similar way, this time using the latest model for MR search. Two trimers were present in the asymmetric unit. Here, the final model contained 193 residues for each of the six monomers, 758 molecules of water and 1 molecule of glycerol.

All refined models showed good agreement with diffraction data, with good geometry and most of the residues in favoured regions of Ramachandran plot (Table 17). In both models a genuine cis-peptide bond was found in the shaft repeat region, involving residues Ala599 and Pro600.

4.4.3 Description of the atomic model

The structure obtained from the MAdV-2 fibre construct 586-787 revealed the presence of a fibre head domain along with two beta-spiral repeats of the fibre shaft domain (Figure 44), although we expected these residues to form the fibre head domain only.

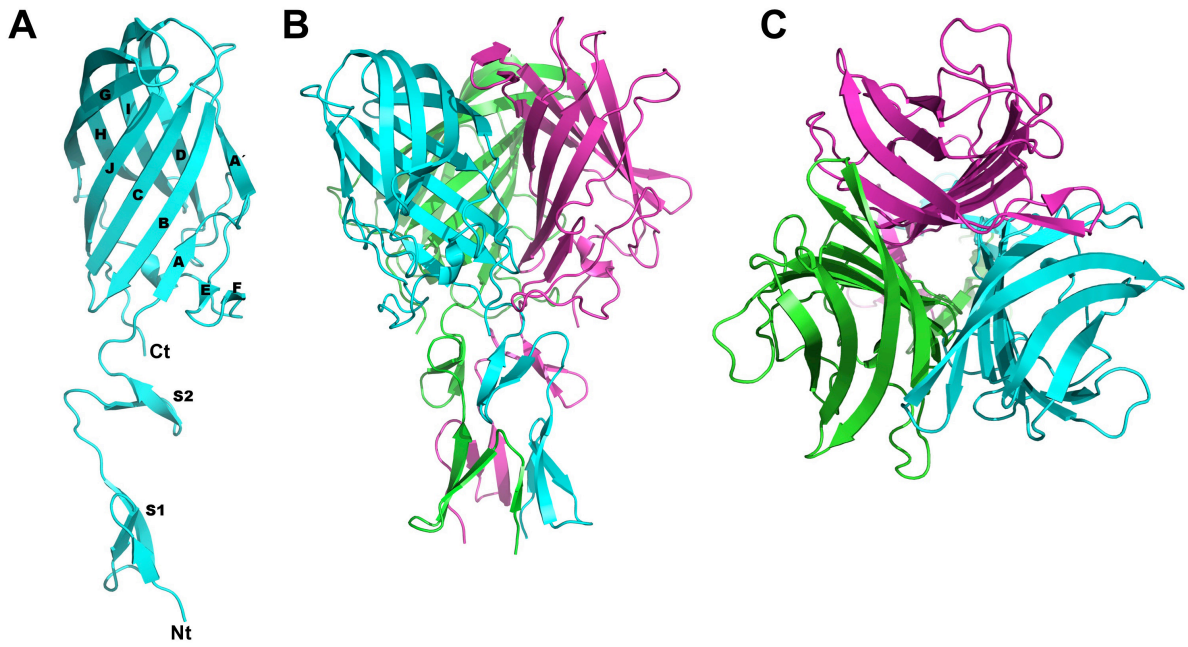


Figure 44: Structure of the MAdV-2 fibre head domain with two shaft repeats. (A) A monomer showing the arrangement of sandwich-like head domain and beta-spiral repeats of the shaft. N and C terminus are indicated while strands are labeled A-J according to the nomenclature proposed by Xia et al. (Xia et al., 1994). Notice the presence of two small strands E and F in the DG loop. beta-spiral repeats are labeled S1 and S2. Side view (B) and top view (C) of the fibre head trimer along with shaft repeats. Between panels A and B the structure has been turned approximately 90º towards the reader. Each monomer is coloured differently.

4.4.3.1 The fibre head domain

Each monomer of the MAdV-2 fibre head contains a beta-sandwich, like other fibre heads for which the structure has been determined. Together they form a trimeric beta-propeller. Each fibre head monomer contains 153 amino acids (634-787), a number less than some other adenovirus fibre heads such as that of HAdV-5 (185 amino acids; Xia et al. (1994)) and CAdV-2 (197 amino acids; Seiradake et al. (2006)).

The fibre head measures 6.1 nm in diameter and has a height of 4.5 nm (Figure 44B). An outwardly flapped HI-loop, among others, especially helps the structure attain a higher diameter than height. Similar flapping-out of the HI-loop can be observed in the HAdV-5 fibre head, which has a diameter of 6.2 nm and a height of 4.0 nm (Figure 45B; Xia et al. (1994)). Both are significantly larger than the previously described SnAdV-1 fibre head,

which is just 3.8 nm tall and has a diameter of 4.6 nm. The anti-parallel beta-sheet of each head monomer can be divided into two halves, FEAA'BCJ and GHID, of which the strand A is kinked. I named this kink A' (Figure 44A). The FEAA'BCJ-sheet faces the inner side of the trimeric assembly and makes several inter-monomeric contacts. On the other hand, the GHID-sheet is mainly solvent exposed. Most of the loops connecting strands are short, with the exception of the DG-loop (residues 690-728). As in many other Mastadenovirus fibre heads, two small beta-strands E and F are present in this loop, between residues 704 and 710. The long DG-loop runs along the bottom of the beta-sandwich and contains amino acid side chains that contribute significantly towards the hydrophobic core of the monomer, as well as making hydrogen bonds and a few hydrophobic contacts with neighbouring monomers (residues 695-700; Figure 44B).

Superposition of monomers of the trimer with each other showed no major conformational deviation, with average r.m.s.d. between 0.6 and 0.7 Å. Differences of more than 1 Å were only observed in parts of the DG-loop that are away from the trimer interface. Similarily, superposition of trimers from the two native models didn't show any difference and they aligned with r.m.s.d. values of less than 0.5 Å.

4.4.3.2 The shaft domain repeats

Beta-spiral repeats are present in adenovirus fibre shaft domains as well as in receptor binding proteins of some other viral fibres such as the mammalian and avian reovirus attachment proteins sigma1 and sigmaC and the bacteriophage PRD1 fibre (Chappell et al., 2002; Guardado Calvo et al., 2005; Merckel et al., 2005). They are characterized by two small beta-strands, connected through a 4 residue loop that has either glycine or proline at the third position (van Raaij et al., 1999b). The presence of glycine or proline facilitates the formation of a type II beta-turn. In the structure, two such glycine-type beta-spiral repeats are present (residues 601-628), and one partial repeat can also be observed (residues 594-598). The only cis-peptide bond of the structure (between residues 599 and 600) is in the 7-residue loop connecting the partial repeat and the following repeat. Amino acid residue in each beta-strand of the repeats varies in number (between 2-5) while the loop connecting two repeats contains 7 residues. Similar glycine-type repeats were observed in HAdV-2 shaft; reovirus attachment proteins also have proline-type repeats (Chappell et al., 2002; Guardado Calvo et al., 2005; van Raaij et al., 1999b). The last shaft repeat is followed by a five amino acid linker mostly consisting of hydrophobic amino acids, which joins the shaft to the C-terminal head domain (629-TSPVF-633). Due to the nature of these amino acids present at shaft-fibre

juncture, little flexibility can be anticipated in MAdV-2 fibre shaft-head junction. The linker present in HAdV-2 consists of six residues that are hydrophilic (393-NKNDDK-398; van Raaij et al. (1999b)).

4.4.3.3 Stability of MAdV-2 fibre

MAdV-2 fibre has a surface area comparable to other mammalian fibre heads. The range of inter-subunit contacts extends from the base of the tail to almost the top of the assembly. These contacts are extensive and include several hydrophobic residues at the shaft as well at the hydrophobic core of the head. The fibre head and shaft jointly have a monomeric surface area of 11 x 10^3 Å^2, of which 17% gets buried upon trimer formation. Considering only the fibre head, these values reduce to 7.9 x 10^3 Å^2 and 17% respectively. The shaft domain, being extensively inter-twinned, buries 43% (1.4 x 10^3 Å^2) of its monomeric surface area, which is 3.4 x 10^3 Å^2. There are 19 inter-monomer hydrogen bonds if both head and shaft are taken together, whereas this value drops to only eight upon removing shaft repeats from the structure. This suggests that the shaft domain contributes significantly in maintaining the integrity of the trimer. Besides that, 3 salt bridges are formed at each monomeric interface, out of which 2 are contributed by shaft repeat residue 615. The calculated energy gain upon trimer formation is around 80 kcal/mol, of which 38 kcal/mol comes from the shaft region. None of the four cysteine residues present in MAdV-2 fibre C-terminal construct is involved in disulphide bond formation.

4.4.4 Structural homologues of the MAdV-2 fibre head

The closest structural homologue of the MAdV-2 fibre head domain turned out to be CAdV-2 fibre head domain (PDB code 2J1K; Seiradake et al. (2006)). When both structures are superposed, they align with an r.m.s.d. of 2.6 Å for 411 C-alpha atoms (Z score 15.8; Figure 45A). The next most similar protein identified is the HAdV-35 fibre head domain (PDB entry 3BQ4) with a r.m.s.d. of 2.7 Å for 412 superimposed C-alpha atoms (Z-score 14.5). Finally, when the structure is superposed onto the mammalian reovirus sigma1 attachment protein (PDB entry 1KKE), the r.m.s.d. is 4.4 Å for 372 superimposed C-alpha atoms (Z-score 8.5).

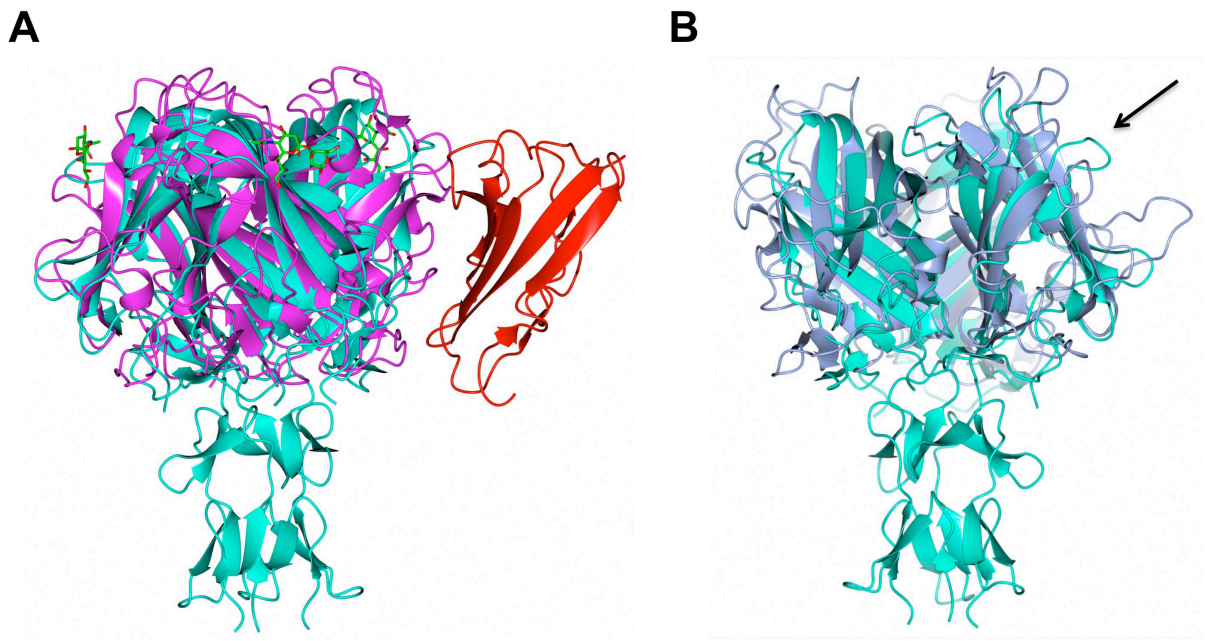


Figure 45: Superposition of MAdV-2 C-terminal construct (cyan) onto CAdV-2 fibre head domain bound to CAR D1 and sialic acid (A; PDB code 2J1K; magenta) and HAdV-5 fibre head domain (B; PDB code 1KNB; blue). One CAR D1 (red) and three sialic acid molecules (green stick) are shown. CAdV-2 fibre head can bind CAR D1 and sialic acid simultaneously on its peripheral region. Both binding sites are close but independent of each other. Black arrow indicates location of the AB loop.

4.4.5 Receptor binding

Surface-exposed loops of the MAdV-2 fibre head have a certain degree of resemblance to loops of other Mastadenovirus fibre heads such as CAdV-2 and HAdV-5 (Figure 45). As mentioned above, CAdV-2 can bind three CAR D1 and three sialic acid molecules simultaneously with its trimeric head (Seiradake et al., 2009). The two binding sites are close to each other but independent. Interaction with CAR D1 is mainly with the AB-loop and partly by the B strand and the DE-loop of each monomer. In contrast to what is observed for HAdV-12 and HAdV-37, the neighbouring monomer does not participate in CAR D1 binding for the CAdV-2 fibre head. None of the residues involved in the interface between the CAdV-2 fibre head and CAR D1 are conserved in the MAdV-2 fibre head, except one (Thr441 in the DE-loop). The AB-loop is also shorter and protrudes somewhat less. However, residue conservation is not considered as a critical criterion for fibre head-CAR D1 interaction (Seiradake et al., 2006). Moreover, human CAR D1 and mouse CAR D1 are 91% identical at sequence level with 17 out of 19 residues that are involved in the interaction between HAdV-12 and CAR D1 being conserved (van Raaij et al., 2000). Based on these facts, I decided to probe binding of MAdV-2 fibre head with D1 domain of CAR by gel-filtration. MAdV-2 fibre and CAR D1 were mixed in a trimer to monomer molar ratio of 1:3,

incubated and analyzed by gel filtration. Both proteins elutes in distinct peaks, showing no evidence for binding to each other (Figure 46).

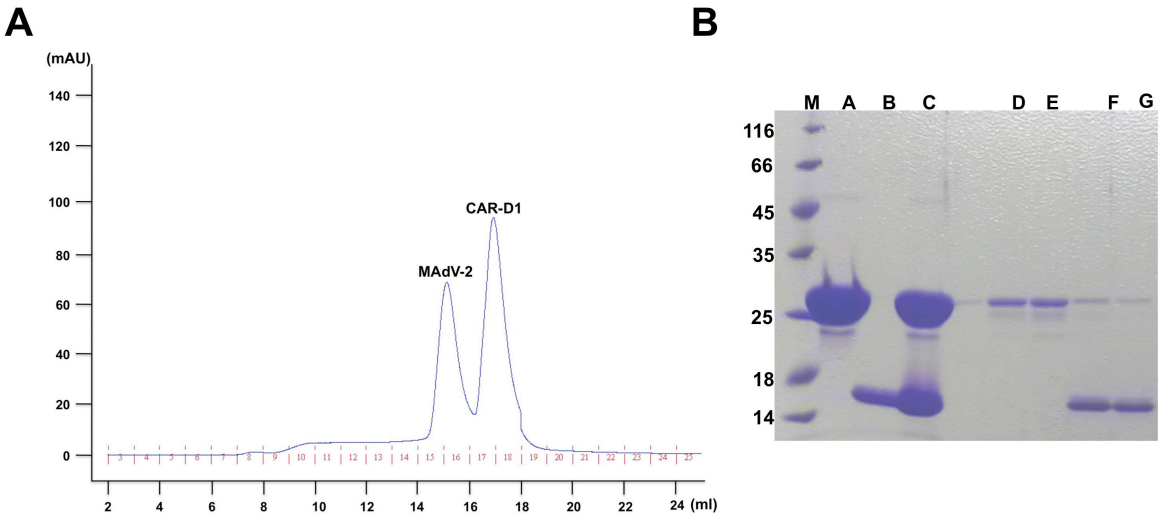


Figure 46: Gel filtration assay of the MAdV-2 fibre head and CAR D1 domain. (A) Elution profile of MAdV-2 fibre head and CAR D1 when mixed in a trimer to monomer molar ratio of 1:3 and passed through superdex 200 column. (B) SDS-gel of pooled fractions and controls. Lanes M: molecular weight marker, A: MAdV-2 fibre head alone (monomeric MW 25 kD), B: CAR D1 alone (MW 15 kD), C: mixture of fibre head and CAR D1 before loading onto column, D and E: collected fractions for MAdV-2 fibre head and F and G: collected fractions for CAR D1. All samples were heated in sample loading buffer at 95 $^{\circ}$C for 5 mins before loading onto SDS-gel.

The surface of MAdV-2 fibre head has a mixed distribution of predicted positively and negatively charge patches (Figure 47). Basic patches can easily be seen on the exposed loop regions, especially on the peripheral areas and on top of the trimer, raising a possibility that they could bind a negatively charged molecule such as sialic acid. The CAdV-2 fibre head engages sialic acid using one of its peripherally located basic patches and its close similarity with the MAdV-2 fibre head structure prompted me to have glycan array experiments performed.

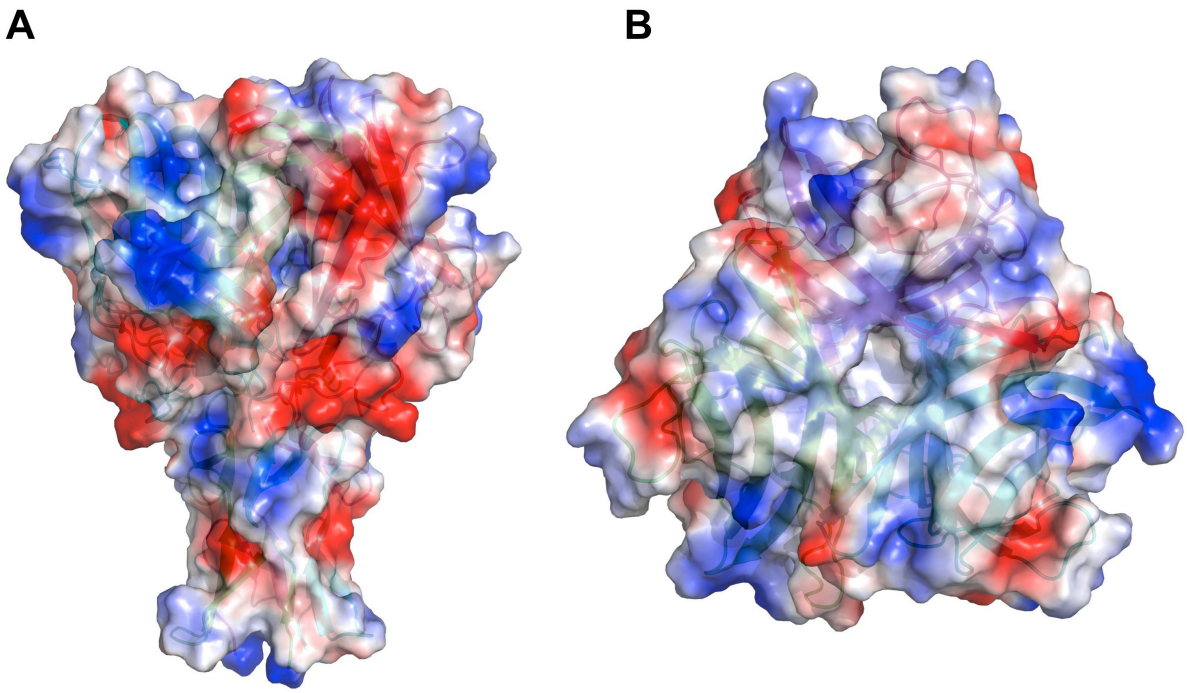


Figure 47: Surace charge distribution of MAdV-2 C-terminal construct 586-787. (A) Side view in the same orientation as figure 44B. (B) Top view in the same orientation as figure 44C.

4.4.6 The MAdV-2 C-terminal construct binds N-acetylglucosamine

N-acetylglucosamine or GlcNAc is a monosaccharide derivative of glucose and is present in a number of glycosaminoglycans such as heparan sulfate and hyaluronan (Esko et al., 2009). Given the wide distribution of glycosaminoglycans in form of proteoglycans in diverse tissue types and extracellular matrix, many pathogenic microorganisms, including viruses, have adopted themselves to successfully use these molecules for attachment and cellular entry (Rostand and Esko, 1997). Glycan array screening of the MAdV-2 C-terminal construct against 74 neoglycoconjugates showed strong binding to glucose derivative GlcNAc (Figure 48). When the glycan array experiment was repeated in presence of 50 mM GlcNAc, partial inhibition in ligand binding was observed indicating that GlcNAc is able to bind MAdV-2 in solution as well. Binding was also observed with BSA conjugated mannose, albeit with lower intensity.

ITC confirmed the interaction of MAdV-2 fibre with GlcNAc where a K_D value of 2 mM was obtained when a model of one binding site per protein monomer subunit was used (Figure 49).

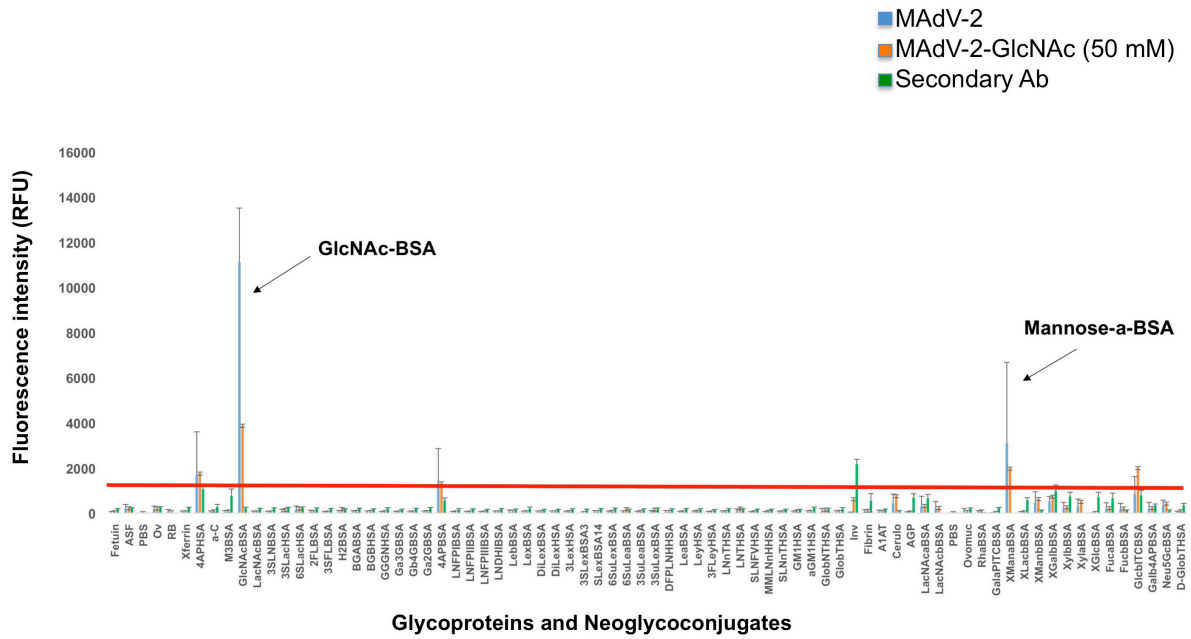


Figure 48: Histogram representing recognition of neoglycoconjugates and glycoproteins by His-tagged MAdV-2 fibre construct 586-787 (20 mg/mL). The most intense signal is detected from glucose derivative carbohydrate GlcNAc, which gets partially inhibited in presence of 50 mM GlcNAc (see legend). (See table AI in annex for details of neoglycoconjugtes and glycoproteins).

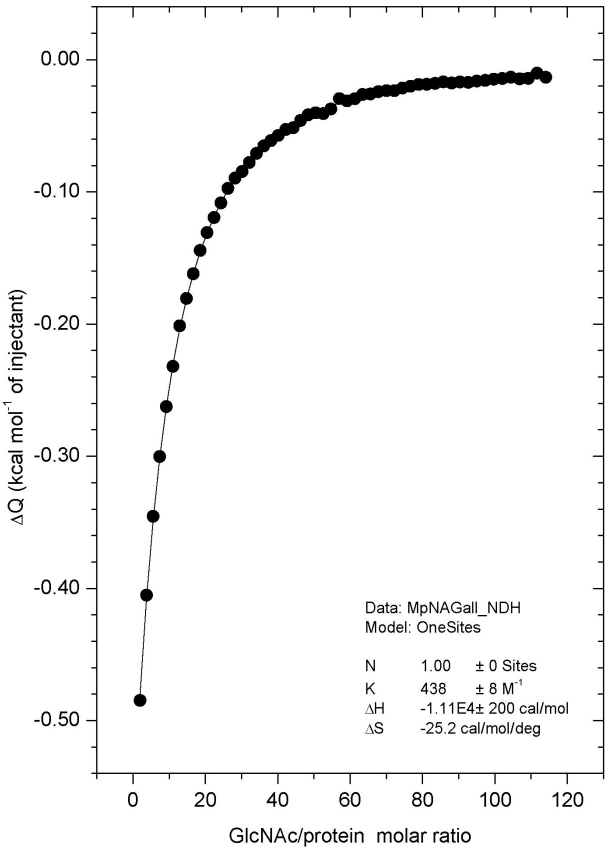


Figure 49: ITC profile of MAdV-2-GlcNAc interaction. Thermodynamic parameters (ΔH and ΔS) and association constant (K) values are indicated.

Involvement of heparan sulphate in HAdV-2 and HAdV-5 infections has been suggested in the past (Dechecchi et al., 2000). It was observed that removal of heparan sulfate-glycosaminoglycans from cells inhibit viral binding and infection. Interestingly, similar abolishment was not observed upon removing chondroitin sulfate glycosaminoglycans, which contain N-acetylgalactosamine as the repeating sugar unit instead of GlcNAc. Based on analyses carried out on a large set of heparin-binding proteins, two major glycosaminoglycan-engaging sequence motifs have been proposed: BBXB and BBBXXB, where B is a basic amino acid (Templeton, 1992). The shaft of HAdV-5 contains a repeat motif KKTX (residue 91-94) and possible involvement of this motif in glycosaminoglycan binding has been postulated (Dechecchi et al., 2000; Dechecchi et al., 2001) and supported by later studies (Bayo-Puxan et al., 2006; Smith et al., 2003; Bayo-Puxan et al., 2009). However, a sequential arrangement of these basic residues may not be required and an epitope formed by conformational closeness of otherwise distantly located basic residues might also be able to bind glycosaminoglycans (Chen et al., 1997). Although it is not clear whether the binding site of heparan sulfate involves GlcNAc or some of its other epitopes, the saccharidic nature of GlcNAc make it one of the most probable candidates to be involved in these interactions.

4.4.7 Conclusion

The structure of a C-terminal fragment of MAdV-2 fibre was determined. The structure revealed the presence of two beta-spiral repeats of the shaft domain along with a somewhat smaller than expected fibre head domain. This is only the second adenovirus fibre for which triple beta-spiral repeats were visualized. The MAdV-2 fibre shaft repeats are of the glycine type. The head domain of the MAdV-2 fibre is structurally very similar to the CAdV-2 fibre head and exhibits charged surface patches. As CAdV-2 can bind CAR D1 and sialic acid, I investigated CAR D1 binding and carbohydrate binding. Glycan array and ITC experiments revealed GlcNAc as a potential ligand. Interaction with human CAR D1 could not be demonstrated.

4.5 Preliminary X-ray crystallographic and biophysical studies of snake adenovirus-1 capsid protein LH3

The Atadenovirus capsid contains a surface exposed, trimeric protein called LH3, which is not present in any other genus of *Adenoviridae*. The function of LH3 is not known, but it may help to stabilize the capsid, be involved in host cell receptor engagement, or both. The SnAdV-1 LH3 has been expressed in bacteria, purified and crystallized. A high-resolution native data set at 2.0 Å was obtained, but phasing attempts were not successful.

4.5.1 Purification and crystallization

4.5.1.1 Construction of the expression vector

A gene fragment coding for full length LH3 protein (373 amino acids; monomeric MW 42 kD) was amplified from SnAdV-1 genomic DNA by PCR and cloned into pET28a(+) bacterial expression vector under the T7 promotor. The protein was expressed with an N-terminal 6xHis purification sequence (M GSSHH HHHHS SGLVP RGSHM ASMTG GQQMG RGSEF).

4.5.1.2 Protein purification and crystallization

The LH3 protein was expressed at low temperature in bacteria. For purification, the protein was subjected to nickel affinity chromatography, followed by strong anion exchange chromatography. In nickel affinity chromatography, the protein eluted in a broad imidazole concentration range (0.1-0.4 M), fractions of which were checked for purity and pooled together to be loaded onto a strong anion exchange chromatography column (the predicted pI of the protein is 5.8). When eluted with a linear gradient of sodium chloride, the protein comes out in a concentration range of 0.3-0.4 M sodium chloride and in a single peak (Figure 50A). The oligomeric nature and purity of the protein was checked by running unheated and heated samples on a SDS-gel, where monomeric protein can only be observed in heated fractions (Figure 50B), while unheated samples show SDS-stable trimers.

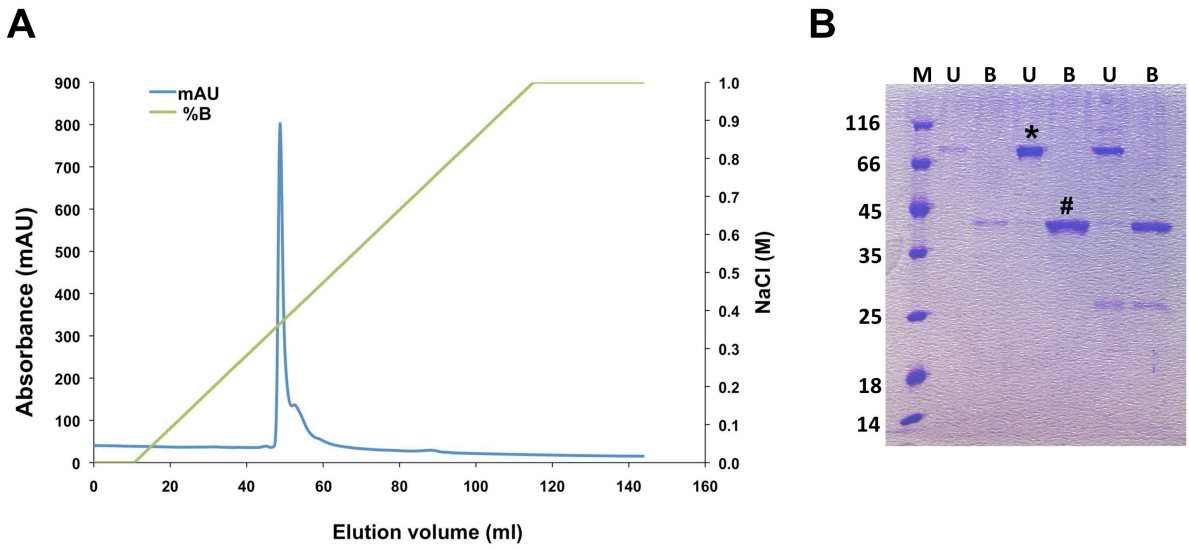


Figure 50: Anion exchange chromatography of LH3 protein. (A) Chromatogram showing elution profile of LH3 when subjected to a linear gradient of sodium chloride. (B) Purity and trimeric nature of LH3 was assessed by denaturing SDS-PAGE, analysing sample both unheated (U) or heated to 95 ºC (B) in denaturing sample buffer before analysis. Fractions collected after anion exchange chromatography. Asterisk indicates the trimeric protein, monomeric protein (#) can be observed in the heated samples.

Extensive crystallization trials were performed, in total testing more than a thousand different conditions. Diffraction-quality crystals were obtained in two conditions, condition 1, which contained 20% (w/v) PEG 3350 and 0.2 M ammonium acetate; and condition 2, which contained 20% (w/v) PEG 3350 and 0.2 M di-ammonium hydrogen phosphate. Crystals grown in condition 1 diffracted X-rays to around 1.7 Å resolution and a complete dataset could be collected to 2.0 Å. Fine, plate like crystals belonged to space group $P2_1$ and apparently contain two protein trimers per asymmetric unit. This would leave 42.9% solvent in the asymmetric unit (Figure 51; see Table 18 for data statistics). A self-rotation function suggested the presence of independent non-crystallograhic three-fold symmetry axes, which supports the presence of two trimers the asymmetric unit of the crystal (Figure 53).

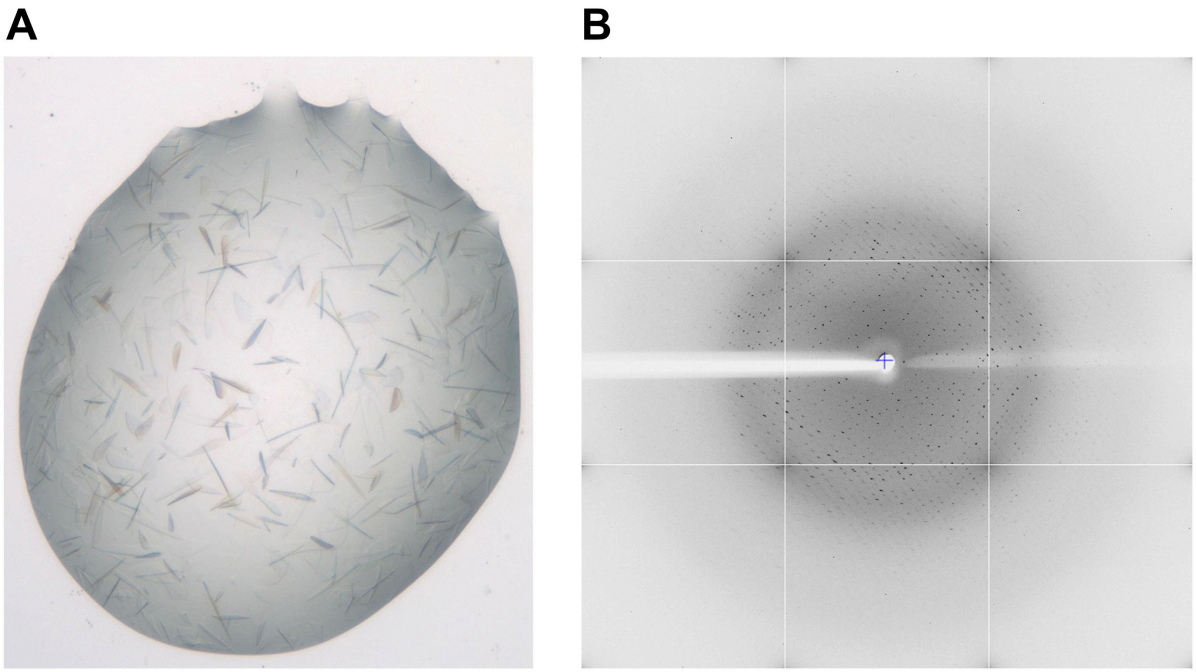


Figure 51: Crystallization and data collection from LH3 crystals. (A) Thin, plate like crystals grew from solutions containing 20% w/v PEG 3350 and 0.2 M ammonium acetate. (B) Image showing diffraction pattern of one of the crystal. Useful data integrated up to 2.0 Å. The beamstop is indicated by a cross-hair mark.

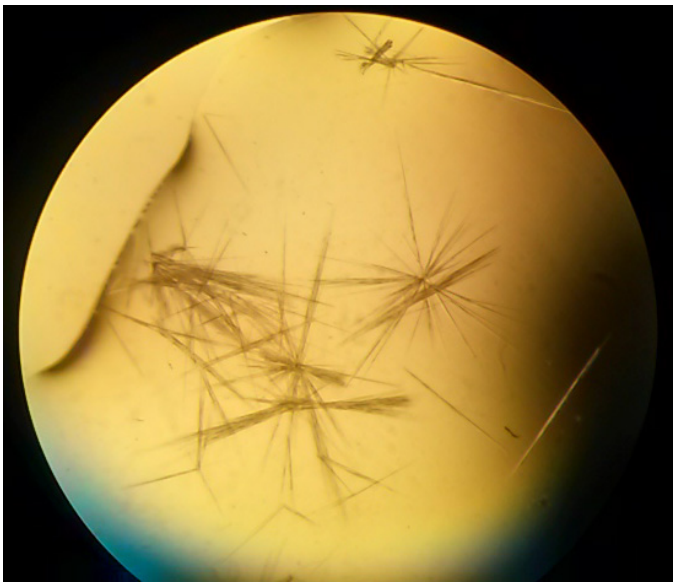


Figure 52: Crystals obtained from methylated LH3 protein.

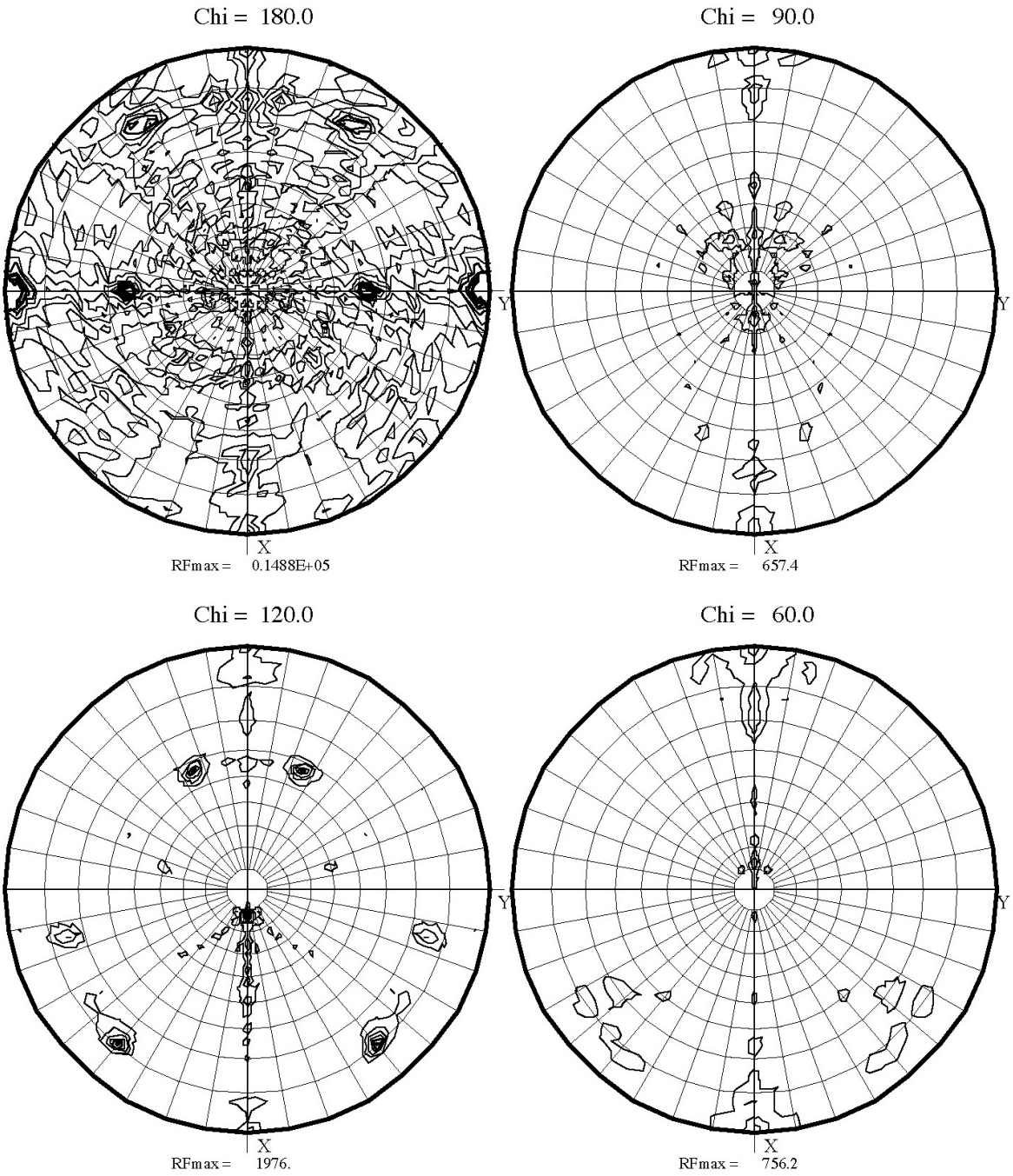


Figure 53: Self-rotation function of the LH3 native crystal data set. Two strong peaks in the chi=120^0 section suggest the presence of two trimeric molecules in the crystal asymmetric unit. Each section is marked on the top with its respective chi value.

Table 18: Crystallographic data measured from LH3 native crystal

Data collection	
Construct (residues)	1-373
Synchrotron, beamline	ESRF, BM30
Detector	ADSC Q315r
Distance (mm)	256.20
Wavelength (Å)	0.9796
Number of images	180
Oscillation range (º)	1.0
Space group	$P2_1$
Cell edges (Å, a, b, c)	70.57, 126.64, 120.61
Resolution range (Å)	67.6-2.0 (2.11-2.0)
Reflections	130096 (17723)
Multiplicity	3.3 (3.3)
Completeness (%)	95.8 (89.6)
Mean <I/s(I)>	3.7 (1.4)
R_{sym} (%)[a]	21.3 (70.3)
V_M ($Å^3Da^{-1}$)‡	2.16
CC Imean	0.962 (0.628)
CC anom	-0.346 (0.073)
Wilson B ($Å^2$)	10.1

Values in parentheses are for the highest resolution bin.

4.5.2 Characterization of expressed and purified LH3 protein by electron microscopy

Purified LH3 was analyzed by negative stain EM to assess its physical appearance and characteristics (experiments performed by Dr. Carmen San Martin, CNB-CSIC, Madrid). As expected, protein trimers were observed (black arrows in figure 54). However, a significant fraction of the molecules were found in higher order oligomeric states and aggregates (coloured arrows in figure 54). The number of constituting trimers in these oligomers was between 5 and 6.

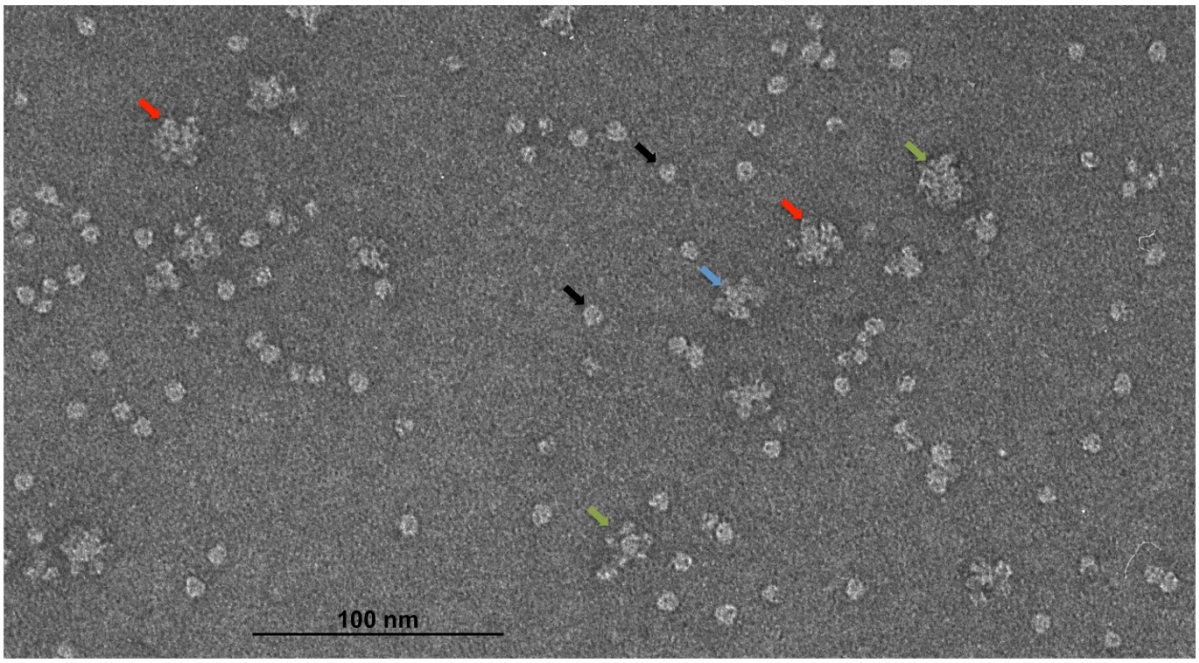


Figure 54: Negative stain EM of purified LH3. Different oligomeric states of the protein can be seen; trimeric protein molecules (black arrows), hexameric particles (red arrows), pentamer (blue arrow) and aggregates (green arrow) are present

4.5.3 Rescue strategies for crystallization

4.5.3.1 Treatment with alpha-chymotrypsin

As mentioned before, LH3 was difficult to crystallize and despite extensive trials, crystals could only be obtained in two conditions. Moreover, these crystals could not be reproduced. Thus, I decided to adopt some rescue strategies. Limited proteolysis experiments were carried out in order to attempt to purify a stable domain of LH3. Prior to proteolysis, samples were incubated at different temperatures to assess their effect on the outcome of proteolysis as described in Material and Methods. I didn't observe any difference on the SDS-gel pattern when samples incubated on three different temperatures (37 $^{\circ}$C, 55 $^{\circ}$C and 65 $^{\circ}$C) and thereafter treated in an identical manner. Thus, I chose 37 $^{\circ}$C as the incubation temperature and performed further experiments with varying ratios of LH3 protein to alpha-chymotrypsin (Figure 55).

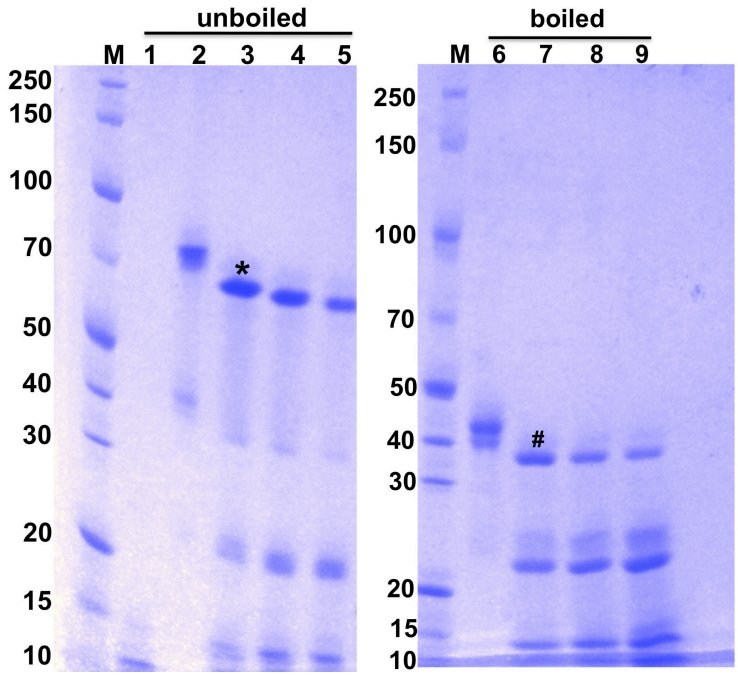


Figure 55: LH3 samples treated with alpha-chymotrypsin at 37 ºC for 20 min and run on a SDS-gel. Lane M: molecular weight marker; Lanes 1-2: alpha-chymotrypsin and untreated LH3 samples respectively as controls; Lanes 3-5: LH3 treated with alpha-chymotrypsin in a sample:enzyme ratio of 1:8, 1:4 and 1:2 respectively. All samples in lanes 1-5 are unheated. Lane 6: untreated LH3 as a control; Lanes 7-9: same as of lanes 3-5. All samples in lanes 6-9 are heated at 95 °C for 5 min prior gel loading. Stable fragment of LH3 retains after treatment is indicated by an asterisk. The hash shows the fragment in monomeric form after heating.

A stable fragment of LH3 was identified. The fragment shows the same pattern of oligomerization as native LH3 and remains intact even after a treatment with relatively high amounts of alpha-chymotrypsin (Figure 55). The fragment along with other digestion products was pooled and purified by anion exchange chromatography (Figure 56A). Although it has been known that α-chymotrypsin has selective preference to cleave aromatic amino acids located at the carboxyl terminal region of proteins (Appel, 1986), western blot analysis was performed to determine the region of proteolytic cleavage. A western blot assay using an anti-his antibody proved the presence of intact N-terminal region in the stable LH3 protein fragment (Figure 56B) hence, I presumed that the proteolytic enzyme has cleaved-off the C-terminal region of the protein. Crystallization trials performed with the purified fragment failed to yield any hit. Crystallization trials with *in situ* proteolysis using the same enzyme were also performed, however, no crystals were obtained either.

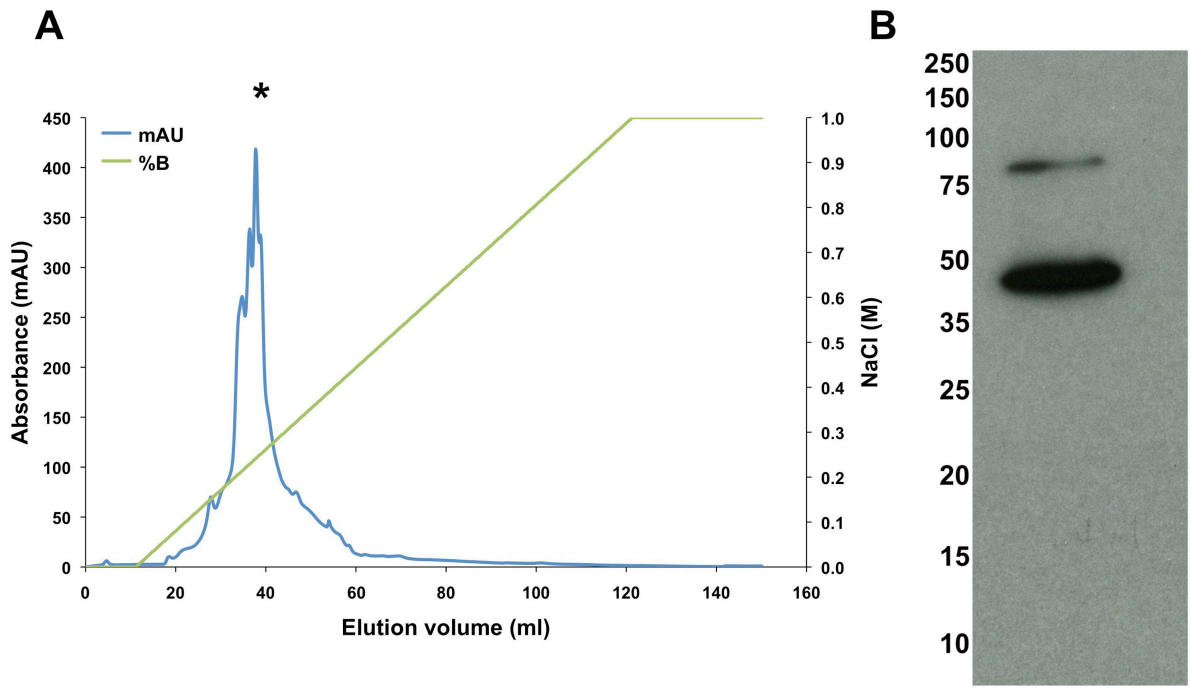


Figure 56: Purification of the LH3 fragment after alpha-chymotrypsin digestion. (A) Anion exchange chromatography profile of fragment obtained after treatment with alpha-chymotrypsin. Asterisk indicates the fraction, which was pooled for western blot analysis and crystallization. (B) Western blot of the pooled fraction. Detection of his tag confirmed the presence of intact N-terminal region in the protein.

4.5.3.2 Methylation of surface exposed lysines

Sometimes side chains of bulky amino acids that are exposed on protein surface hinder the process of crystallization by disturbing crystals lattice contacts. Changing the surface property by modifying such residues could help improving crystal quality as well as increases the chance of getting crystal hits. Reductive methylation of surface exposed lysins is one such method, which has been shown to rescue crystallization when obtaining crystals were difficult otherwise (Walter et al., 2006; Kobayashi et al., 1999). I adopted the protocol as decribed by Walter et al. to assess the feasibility of this method in this case. After a period of 3 months conditions close to the one which produced native LH3 crystals, resulted in morphologically distinct, fine, needle-shaped crystals when drops where setup with methylated version of LH3 (Figure 52). Diffraction of these crystals has not been tested yet.

4.5.4 Binding with carbohydrates

Glycan array experiments were performed with 74 different neoglycoconjugates and glycoproteins. No binding to any of the carbohydrates was evident; the obtained fluroscence signals were all below the experimentally determined threshold value of 220 RFU.

4.5.5 Conclusion

Crystals of the SnAdV-1 LH3 protein were of high quality, and diffracted synchrotron radiation up to a maximum resolution of 2 Å. Analysis of the crystallographic data suggested that the protein is trimeric in nature, as was expected. Due to high reluctancy of the protein to form crystals, structure determination has not been possible so far. I have confirmed that the protein is highly stable towards denaturing agents and that its monomer can only be visualized upon boiling the sample prior gel run. Further biochemical and biophysical analysis indicated that the protein might form higher order oligomers and aggregates. Glycan screening data showed that LH3 has no affinity for carbohydrates, however, considering its localization over the viral capsid, a possible involvement in host-cell recognition can not be completely ruled out.

Among adenoviruses, atadenoviruses are unique in their capsid composition with the presence of two distinct proteins, LH3 and p32k. These ‘non-standard’ proteins may confer on them some distinct functions that have not been previously observed in other adenoviruses or they may be involved in some standard functions such as receptor recognition that are thought to be associated with other, well established proteins. In any case, a deeper and thorough characterization of these proteins, including LH3, will help reveal new facets of these viruses.

Conclusions

5. Conclusions

1. The snake adenovirus 1 fibre protein consists of 345 amino acids instead of 415 as reported in the database.

2. The 1.35 Å structure of snake adenovirus 1 fibre head, the first for an Atadenovirus protein, retains the overall topology of other adenovirus fibre heads, but in a much more compact form.

3. The lizard adenovirus 2 short fibre head structure is virtually identical to that of the snake adenovirus 1 fibre head.

4. The 2.2 Å structure of the turkey adenovirus 3 fibre head, the first for a Siadenovirus protein, revealed a strongly positively charged surface, which may be important for host cell interaction. A beta-hairpin insertion in the C-strand contacting a neighboring monomer is unique to the TAdV-3 structure. Its closest structural homologues were identified as reovirus attachment proteins.

5. Turkey adenovirus 3 fibre head binds 2,3'- and 2,6'-sialyllactose. They bind to the side of the fibre head, involving residues Glu392, Thr419, Val420, Lys421, Asn422, and Gly423.

6. The 1.8 Å structure of the carboxy terminal domain of the mouse adenovirus 2 fibre revealed a beta-sandwich head domain architecture similar to other Mastadenovirus fibre heads. The shaft domain of MAdV-2 is made up of glycine type beta-spiral repeats, a feature also observed in HAdV-2 fibre.

7. *In vitro*, the carboxy-terminal domain of the mouse adenovirus 2 fibre binds N-acetyl-glucosamine.

Sumario en Castellano

6. Sumario en Castellano

Caracterización estructural y funcional de proteínas de la cápsida de adenovirus no humanos

6.1 Resumen

Aunque los adenovirus son agentes causantes de enfermedades, también se utilizan como modelo para investigar los mecanismos básicos de la biología molecular. El uso de vectores basados total o parcialmente en adenovirus no humanos como vectores para ensayos de terapia génica y anticancerígena presenta un interés considerable, ya que permitiría evitar la inmunidad preexistente en humanos, y presentaría la posibilidad de dirigirse a diferentes dianas celulares.

La cápside icosaédrica de adenovirus, además de proteger el ADN de doble cadena, contiene elementos involucrados en la entrada del virus. Uno de estos elementos es el homotrímero de la fibra, la proteína responsable de la interacción inicial virus-hospedador, y por tanto un factor crucial en la biología de los adenovirus y en la exploración de sus aplicaciones médicas. En esta tesis doctoral, se han determinado las estructuras del dominio distal (cabeza) de las fibras de cuatro adenovirus no humanos. Tres de las estructuras pertenecen a géneros recientemente establecidos de la familia *Adenoviridae*: *Siadenovirus* y *Atadenovirus*. Aunque pertenecen a distintos géneros y se dirigen a distintos organismos, las estructuras presentan una topología conservada de sándwich beta en su parte central. Sin embargo, muchas características como el tamaño, la forma y la longitud de los bucles conectores muestran una gran diversidad, lo que podría producir un profundo impacto en el tropismo viral.

En el caso de los *Atadenovirus*, las dos estructuras determinadas (fibra del adenovirus de serpiente tipo 1 y fibra 1 del adenovirus de lagarto tipo 2) muestran que la cabeza de las fibras en este géneroson las más pequeñas de todas las estructuras de fibras de adenovirus resueltas hasta el momento. Esto se debe a que presentan bucles relativamente cortos conectando sus cadenas beta, lo que los hace estructuras más compactas. La presencia en su estructura de una prominente hélice alfa es también una característica destacable encontrada solamente en este género.

El adenovirus de pavo tipo 3, un siadenovirus que existe tanto en forma avirulenta como virulenta, tiene una cabeza de la fibra más similar a las proteínas de unión al receptor de

reovirus y bacteriófagos que a las cabezas de las fibras de otros adenovirus. La presencia de una única horquilla beta hace esta estructura claramente distinguible entre otras proteínas análogas de adenovirus. Además, la notable similitud entre la cabeza de la fibra del adenovirus de pavo 3 y las proteínas de unión al receptor de reovirus ha reforzado la teoría de un vínculo evolutivo entre adenovirus y reovirus. La caracterización estructural de esta cabeza de fibra se ha desarrollado en más profundidad para buscar alguna perspectiva funcional por medio de herramientas biofísicas. Se han identificado como ligandos carbohidratos que contenían ácido siálico, y se han determinado con éxito estructuras de co-cristales proteína-ligando. La estructura de la cabeza de la fibra del adenovirus de pavo 3 unida al ligando ha revelado un sitio completamente nuevo de unión a ácido siálico para las fibras de adenovirus, que se localiza muy cerca de la superficie intermonomérica del trímero.

El adenovirus murino de tipo 2 pertenece al mismo género que los adenovirus humanos (*Mastadenovirus*). Este virus podría utilizarse como sistema modelo para estudiar la patogenicidad de los adenovirus, por lo cual sería de gran ayuda el disponer de la caracterización tanto estructural como funcional de su fibra. En esta tesis se ha resuelto la estructura cristalina del dominio cabeza de la fibra, junto con dos repeticiones precedentes del dominio asta, a alta resolución. Aunque la estructura muestra una alta similitud con otras cabezas de fibras del género *Mastadenovirus*, existen ligeras variaciones que podrían tener implicaciones funcionales. Además, se ha determinado que la N-acetilglucosamina, un componente común de glicosaminoglicanos como el heparán sulfato y que tiene una amplia distribución en diversos tejidos, es un ligando para la cabeza de la fibra del adenovirus murino 2.

En resumen, se presentan nuevos datos estructurales y funcionales acerca de distintas cabezas de fibras de adenovirus, que serán útiles en la elucidación de su tropismo y biología. Las nuevas estructuras también podrían tener utilidad para el diseño y el desarrollo de aplicaciones basadas en adenovirus no humanos.

6.2 Introducción

Los adenovirus se aislaron por primera vez en 1953 de amígdalas faríngeas humanas (Rowe et al., 1953), y posteriormente de otros tejidos adenoideos (Hilleman, 1954; Huebner, 1954). Basándose en estas observaciones de un origen común y el cercano parentesco de los agentes aislados, se les dio el nombre colectivo de adenovirus en 1956 (Enders, 1956).

Durante las dos últimas décadas, los adenovirus han sido investigados con dos objetivos principales: desarrollar antivirales efectivos contra ellos, y utilizar el propio virus como vector para terapia génica y anticancerígena. Además, algunos adenovirus humanos se han usado como vacunas orales contra enfermedades respiratorias.

Aunque los adenovirus se asocian con ciertas condiciones patogénicas, estas condiciones suelen ser autolimitantes y la infección habitualmente induce una fuerte inmunidad adaptativa y protectora. Sin embargo, en el caso de pacientes inmunosuprimidos, la infección puede acarrear complicaciones clínicas severas (Kinchington et al., 2005). A pesar de todos los esfuerzos, actualmente no existe ningún antiviral contra adenovirus por vía sistémica o tópica (Kinchington et al., 2005).

En lo que concierne al desarrollo de vectores, durante la década de los 90 se realizaron esfuerzos significativos en el diseño de vectores basados en adenovirus, debido a sus características favorables: alta capacidad para albergar transgenes, producción de recombinantes con título alto, y transducción eficiente (McConnell and Imperiale, 2004; Bachtarzi et al., 2008). Un inconveniente al usar sistemas de vectores basados en adenovirus humanos, como HAdV-5, es que los pacientes pueden tener inmunidad preestablecida contra ellos. El uso de adenovirus no humanos puede ser ventajoso, ya que podrían ser menos inmunogénicos y tener propiedades innovadoras de unión a receptores (Loser et al., 2002). Por esta razón es necesario caracterizar estos virus, poniendo énfasis en los componentes asociados con el tropismo.

La familia *Adenoviridae* consta de cinco géneros: *Mastadenovirus* (que infectan mamíferos), *Aviadenovirus* (que infectan aves), *Atadenovirus* (que infectan rumiantes, reptiles y aves) , *Siadenovirus* (que infectan aves y ranas) e *Ichtadenovirus* (que infectan peces) (Harrach, 2011). Genéticamente, todos los adenovirus se componen de ADN bicatenario lineal y no segmentado, con distintivas repeticiones terminales invertidas (RTI) y una proteína terminal unida covalentemente al extremo 5' de cada cadena. El tamaño del genoma varía entre 26 y 48 kpb de longitud entre diferentes especies, junto con otras características como la longitud de las RTI y el grado de conservación de la secuencia. Sin

embargo, estas variaciones son mayoritariamente observadas en las regiones terminales, mientras que la región central del genoma suele estar conservada (Figura 2; Mautner et al. (1995); (Harrach, 2011)).

El virión del adenovirus carace de componentes lipídicos (membranas); tiene 90 nm de diámetro y una simetría icosaédrica con un número de triangulación T=25. La partícula madura se compone de una cápside externa y un núcleo interno, que contiene ADN y suman una masa total de 150 x 10^6 Da aproximadamente (Shenk, 1996; Vellinga et al., 2005). Las proteínas aportan el 87% de la masa total, mientras que una única molécula de ADN aporta el 13% restante. Las caras de la cápside del adenovirus están formadas por hexones, y los vértices por pentones donde se inserta una fibra trimérica. Además de estas, hay algunas proteínas estructurales minoritarias que son necesarias para el ensamblaje y que funcionan como estabilizadoras de la cápside (San Martin, 2012). En los *Atadenovirus*, la cápside contiene dos proteínas específicas de género, p32 y LH3, que se ha propuesto que dotan de características distintivas a las especies de este género (Löser, 1999).

El adenovirus de serpiente tipo 1 (SnAdV-1) fue aislado por primera vez de una serpiente del maíz (*Elaphe guttata*), que mostraba síntomas clínicos de neumonía (Juhasz and Ahne, 1993). Basándose en la organización y las propiedades filogenéticas de su genoma de 27,7 kpb, fue asignado al género *Atadenovirus* (Harrach, 2011). El gen de la fibra del SnAdV-1 que encontramos en la base de datos (GenBank DQ1064141.1) codificaría un polipéptido de 415 residuos aminoacídicos (Farkas et al., 2008). Sin embargo, en esta tesis doctoral, se ha encontrado que solo tiene 345 aminoácidos de longitud (Singh et al., 2013).

El adenovirus de pavo de tipo 3 (TAdV-3) pertenece al género *Siadenovirus* y presenta dos formas distintas: la forma virulenta, que causa enfermedades a menudo letales tal como la enteritis hemorrágica en pavos y la enfermedad del bazo marmóreo en faisanes (Domermuth et al., 1979; Palya et al., 2007); y la forma avirulenta o menos virulenta que se utiliza como vacuna contra la enteritis hemorrágica del pavo (Domermuth et al., 1977; Silim et al., 1978; Beach et al., 2009b). La comparación de las secuencias genómicas de las cepas virulentas y avirulentas de TAdV-3 mostrómutaciones en la secuencia de la fibra (Beach et al., 2009a).

La proteína de la fibra del TAdV-3 contiene 454 residuos, y se ha predicho que está formada por un dominio cola (residuos 1-45), un dominio asta con repeticiones putativas en triple espiral beta entre los residuos 46 y 303, y un dominio cabeza que abarcaría la parte restante (aminoácidos 304 al 454). Las mutaciones antes mencionadas se localizan en el

dominio cabeza (residuos 354 y 376), y consisten en una sustitución de metioninas en la forma avirulenta por isoleucina y treonina, respectivamente, en la cepa virulenta.

El adenovirus de lagarto de tipo 2 (LAdV-2) es el primer *Atadenovirus* identificado hasta el momento que posee dos tipos de fibra distintos en su virión (Penzes et al 2014). El gen de la fibra 1 de LAdV-2 codifica una proteína de 331 aminoácidos con un peso molecular de 34,7 KDa, mientras el producto del gen de la fibra 2 tiene 433 aminoácidos con un peso molecular de 45,7 kDa. El modelo predicho para la fibra 1 muestra un dominio cola del residuo 1 al 40, mientras que el dominio asta empieza alrededor del residuo 45 y continúa hasta el residuo 210. Posteriormente, tras una corta región conectora, se encontraría el dominio cabeza. La fibra 2 es 102 aminoácidos más larga que la fibra 1, y su dominio cola predicho va del aminoácido 1 al 45. La primera repetición en espiral beta del asta se observa alrededor del residuo 46, mientras el extremo carboxilo terminal a partir del residuo 261 de la proteína forma el dominio cabeza.

Muy recientemente, se ha realizado la caracterización del genoma completo del adenovirus murino de tipo 2 (MAdV-2) (Hemmi et al., 2011), el más largo de los tres tipos de adenovirus murinos secuenciados. Se encontró un único gen de fibra, que codifica una proteína con una longitud de 787 aminoácidos. La fibra del MAdV-2 es la más larga comparada con las fibras del MAdV-1 y el MAdV-3 y también con las fibras del HAdV-2, 3 y 5. La predicción de estructura secundaria de la fibra confirma su organización general que coincide con las otras fibras de adenovirus. La conformación predicha seria un dominio cola corto (residuos 1-50), un dominio asta central con repeticiones en espiral beta (residuos 53-630) y un dominio cabeza distal en el extremo carboxilo terminal (residuos 633-787). Justo antes del final del dominio cola, entre los residuos 40-47, puede apreciarse una región corta rica en prolina que probablemente ayuda a la fibra a unirse a la base de pentón sobre la cápside.

6.3 Objetivos

El principal objetivo de esta tesis es la caracterización estructural y funcional de las proteínas externas de la cápside de varios adenovirus no humanos: SnAdV-1, TAdV-3, LAdV-2 y MAdV-2. Las estructuras de estas proteínas podrían dar pistas sobre la unión a sus receptores y conducir a aplicaciones relacionadas con terapia génica y anticancerígena basada en adenovirus. Los objectivos concretos que se han planteado han sido los siguientes:

1. Clonación, expresión, purificación y cristalización de la proteína de la fibra del SnAdV-1. Resolución de la estructura del dominio cabeza y su comparación con otras estructuras de fibras de adenovirus conocidas. Identificación de posibles ligandos.

2. Clonación, expresión, purificación y cristalización de la proteína de las fibras virulenta y no virulenta del TAdV-3. Resolución de la estructura del dominio cabeza. Comparación de las estructuras entre sí y con otras estructuras de fibras de adenovirus conocidas. Identificación de posibles ligandos, determinación de la afinidad del ligando con la fibra y resolución de la estructura del dominio cabeza con su ligando.

3. Clonación, expresión, purificación y cristalización de las proteínas de las fibras tanto larga como corta del adenovirus de lagarto 2. Resolución de la estructura del dominio cabeza de unión al receptor. Comparación de la estructura con la del dominio cabeza de la fibra del adenovirus de serpiente 1.

4. Clonación, expresión, purificación y cristalización de la proteína de la fibra del MAdV-2. Resolución de la estructura del dominio carboxi-terminal. Comparación de la estructura con otras estructuras de fibras de adenovirus conocidas. Identificación de posibles ligandos.

5. Clonación, expresión, purificación y cristalización la proteína minoritaria específica de atadenovirus LH3 del SnAdV-1.

6.4 Resultados y discusión

6.4.1 Estructura del dominio cabeza de la fibra de adenovirus de serpiente tipo 1

El adenovirus de serpiente tipo 1 es el primer *Atadenovirus* del que se ha resuelto la estructura de la cabeza de la fibra. La estructura trimérica compacta es muy estable y conlleva interacciones no covalentes extensivas entre sus unidades monoméricas. Desde el punto de vista de las dimensiones y de longitud, es la cabeza de fibra de adenovirus más pequeña conocida hasta el momento, junto con la cabeza de la fibra 1 del *atadenovirus* de lagarto tipo 2. Sin embargo, mantiene la misma topología en sándwich beta ABCJ-GHID que las otras fibras. Es interesante ver que la estructura es más similar a las fibras de los bacteriófagos que a otras cabezas de fibras de adenovirus. Por medio de una criba con matriz de glicanos (glycan array) se identificó a la sialil-lactosa como un ligando potencial. Esto sugiere que el virus puede utilizar un glicano de la superficie celular como receptor, sin embargo puede que dicho glicano sea más complejo que los analizados.

6.4.2 Estructura y unión de sialil-lactosa al dominio cabeza de la fibra de un Siadenovirus, el adenovirus de pavo 3

Se han resuelto las estructuras atómicas de la cabeza de la fibra de ambas cepas del TAdV-3: la avirulenta y la virulenta. Este es el primer *Siadenovirus* del que se ha descrito la estructura de la cabeza de la fibra. La estructura reveló una superficie con fuerte carga positiva, que podría ser importante para la interacción con su célula hospedadora. Se han demostrado las interacciones de la cabeza de la fibra con glicanos sialidados en ambos tipos de virus, y se han determinado estructuras cristalográficas de la fibra con el ligando. Las estructuras cristalográficas de estos complejos revelaron el sitio de unión al ácido siálico de la cabeza de la fibra del TAdV-3, que se localiza en la superficie intermonomérica y está constituido por lazos expuestos en la superficie. Una inserción de horquilla beta que interacciona con un monómero vecino es única entre las fibras de adenovirus. La falta de similitud de secuencia con otras cabezas de fibras de *Siadenovirus* no permite predecir si otras fibras de Siadenovirus también contendrían esta horquilla beta, aunque es posible. La cabeza de la fibra de TAdV-3 comparte su topología en barril beta ABCJGHID con cabezas de fibras de reovirus, mientras que las fibras de bacteriófagos y otras fibras de adenovirus tienen un sandwich beta con láminas ABCJ y GHID separadas. Los reovirus son virus con ARN de doble cadena, y por tanto no se prevén relaciones evolutivas directas entre adenovirus y reovirus. Por esta razón no esta claro cómo los genes de la fibra pueden haberse

transferido entre ellos, o si el proceso evolutivo implicado ha sido convergente en lugar de divergente. El conocimiento de la estructura puede permitir también el diseño de adenovirus quiméricos que incorporen la cabeza de la fibra del TAdV-3 e investigar su tropismo, que a su vez puede conducir a vacunas o vectores de terapia génica que puedan dirigirse a tipos celulares específicos.

6.4.3 Estructura cristalina de la cabeza de la fibra corta del adenovirus de lagarto tipo 2

Se ha determinado la estructura cristalográfica de la cabeza de la fibra corta del LAdV-2 refinada hasta una resolución máxima de 2 Å. Es la segunda proteína de *Atadenovirus* de la cual se ha determinado la estructura, siendo la primera la cabeza de la fibra de SnAdV-1 descrita en un apartado anterior. La estructura es muy similar a la de la cabeza de la fibra de SnAdV-1, lo cual era de esperar debido a una similitud del 54% en la secuencia. También se intentó cristalizar la cabeza de la fibra larga, que sólo tiene un 12% de similitud de secuencia con la cabeza de la fibra corta, pero no se obtuvieron criatales que difractaran adecuadamente. Estudios futuros estarán dirigidos a obtener informacion de las propiedades de unión al receptor de estas proteínas.

6.4.4 Estructura cristalina de la cabeza de la fibra del adenovirus murino 2 con las repeticiones del asta adyacentes

La estructura del fragmento C-terminal de la fibra del MAdV-2 reveló la presencia de dos repeticiones en espiral beta en el dominio asta, junto con un dominio cabeza más pequeño de lo esperado. Esta es la segunda fibra de adenovirus en la que se han resuelto repeticiones en espiral triple beta. Como en el caso de la fibra del HAdV-2, las repeticiones del asta de la fibra del MAdV-2 están formadas por glicinas con estructura en forma de cadena beta de longitud variable. El dominio cabeza de la fibra del MAdV-2 es muy similar a la cabeza de la fibra de CAdV-2 y muestra partes cargadas en su superficie. Como CAdV-2 puede unir CAR D1 y ácido siálico, se investigó la unión a CAR D1 y a carbohidratos. Mediante ensayos de unión a una matriz de glicanos se observó que el GlcNAc podría ser un ligando potencial; sin embargo, no fue posible demostrar la interacción con el dominio D1 de la proteína CAR humana.

6.4.5 Estudios preliminares por cristalografía de rayos X y estudios biofísicos de la proteína LH3 de la cápsida de adenovirus de serpiente 1.

Se han realizado estudios preliminares por cristalografía de rayos X y estudios biofísicos en la proteína LH3, específica del género *Atadenovirus*. Se obtuvieron cristales nativos de gran calidad que difractaron a 2 Å, y su análisis sugirió que la proteína era trimérica. Debido a la dificultad de obtener cristales de la proteína, no ha sido posible hasta el momento determinar su estructura. Se ha confirmado que la proteína es altamente estable frente a agentes desnaturalizares, y que su monómero sólo puede visualizarse tras hervir la muestra antes de someterla a electroforesis desnaturalizante en gel de acrilamida. Análisis bioquímicos y biofísicos adicionales indicaron que la proteína podría formar agregados y oligómeros de orden superior. Los datos de cribado de glicanos mostraron que LH3 no tiene afinidad por carbohidratos; sin embargo, debido a su localización en la cápside viral, no puede descartarse que esté involucrado en el reconocimiento de la célula hospedadora.

La proteína LH3, junto con otra denominada p32k, forma el conjunto de proteínas minoritarias de la cápsida específicas de *Atadenovirus*. Estas proteínas podrían conferirles distintas funciones que no se han observado previamente en otros adenovirus. Además, LH3 por su posición externa en la cápsida, podría estar involucrada en funciones de reconocimiento del receptor, asociándose con otras proteínas bien establecidas. En cualquier caso, una caracterización más profunda de estas proteínas ayudará a revelar nuevas facetas de estos virus.

6.5 Conclusiones

1. La proteína de la fibra del SnAdV-1 consta de 345 aminoácidos en lugar de los 415 registrados en la base de datos.

2. La estructura de la cabeza de la fibra del adenovirus de serpiente 1, resuelta a 1.35 Å de resolución, es la primera descrita dentro del género *Atadenovirus* y conserva la topología de otras cabezas de fibra de adenovirus, pero es mucho más compacta.

3. La estructura del dominio de cabeza de la fibra corta del adenovirus de lagarto, resuelta a 2.0 Å de resolución, es muy similar a la del adenovirus de serpiente.

4. La estructura del dominio cabeza del adenovirus de pavo 3 resuelta a 2.2 Å es la primera resuelta dentro del género *Siadenovirus*. Posee una superficie cargada positivamente, lo que probablemente es importante para la interacción con la célula hospedadora. Una inserción en forma de horquilla beta que contacta con un monómero vecino es una característica única de la estructura de cabeza de la fibra del adenovirus de pavo 3. Sus homólogos estructurales más próximos han sido identificados como proteínas de unión al receptor de reovirus.

5. El dominio de cabeza de la fibra del adenovirus de pavo une 2,3'- y 2,6'-sialillactosa. El sitio de unión se compone de los residuos Glu392, Thr419, Val420, Lys421, Asn422, y Gly423.

6. La estructura del dominio de cabeza de la fibra del adenovirus de ratón 2 resuelta a 1.8 Å, reveló una estructura en sandwich beta muy similar a la de otros Mastadenovirus. El dominio asta de la estructura resuelta está compuesto de dos repeticiones en espiral beta de tipo glicina.

7. *In vitro*, el dominio carboxi-terminal de la fibra del adenovirus de ratón une N-acetilglucosamina.

References

7. References

Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr D Biol Crystallogr *52*, 30-42.

Albinsson, B., and Kidd, A.H. (1999). Adenovirus type 41 lacks an RGD alpha(v)-integrin binding motif on the penton base and undergoes delayed uptake in A549 cells. Virus Res *64*, 125-136.

Anderson, C.W. (1990). The proteinase polypeptide of adenovirus serotype 2 virions. Virology *177*, 259-272.

Anderson, C.W., Young, M.E., and Flint, S.J. (1989). Characterization of the adenovirus 2 virion protein, mu. Virology *172*, 506-512.

Appel, W. (1986). Chymotrypsin: molecular and catalytic properties. Clin Biochem *19*, 317-322.

Arnberg, N., Edlund, K., Kidd, A.H., and Wadell, G. (2000). Adenovirus type 37 uses sialic acid as a cellular receptor. J Virol *74*, 42-48.

Arnberg, N., Kidd, A.H., Edlund, K., Nilsson, J., Pring-Akerblom, P., and Wadell, G. (2002). Adenovirus type 37 binds to cell surface sialic acid through a charge-dependent interaction. Virology *302*, 33-43.

Arnberg, N., Mei, Y., and Wadell, G. (1997). Fiber genes of adenoviruses with tropism for the eye and the genital tract. Virology *227*, 239-244.

Bachtarzi, H., Stevenson, M., and Fisher, K. (2008). Cancer gene therapy with targeted adenoviruses. Expert Opin Drug Deliv *5*, 1231-1240.

Bangari, D.S., and Mittal, S.K. (2005). Porcine adenovirus serotype 3 internalization is independent of CAR and alphavbeta3 or alphavbeta5 integrin. Virology *332*, 157-166.

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr *67*, 271-281.

Bayo-Puxan, N., Cascallo, M., Gros, A., Huch, M., Fillat, C., and Alemany, R. (2006). Role of the putative heparan sulfate glycosaminoglycan-binding site of the adenovirus type 5 fiber shaft on liver detargeting and knob-mediated retargeting. J Gen Virol *87*, 2487-2495.

Bayo-Puxan, N., Gimenez-Alejandre, M., Lavilla-Alonso, S., Gros, A., Cascallo, M., Hemminki, A., and Alemany, R. (2009). Replacement of adenovirus type 5 fiber shaft heparan sulfate proteoglycan-binding domain with RGD for improved tumor infectivity and targeting. Hum Gene Ther *20*, 1214-1221.

Beach, N.M., Duncan, R.B., Larsen, C.T., Meng, X.J., Sriranganathan, N., and Pierson, F.W. (2009a). Comparison of 12 turkey hemorrhagic enteritis virus isolates allows prediction of genetic factors affecting virulence. J Gen Virol *90*, 1978-1985.

Beach, N.M., Duncan, R.B., Larsen, C.T., Meng, X.J., Sriranganathan, N., and Pierson, F.W. (2009b). Persistent infection of turkeys with an avirulent strain of turkey hemorrhagic enteritis virus. Avian Dis *53*, 370-375.

Bell, C.E., and Lewis, M. (2000). A closer view of the conformation of the Lac repressor bound to operator. Nat Struct Biol *7*, 209-214.

Benko, M., and Harrach, B. (1998). A proposal for a new (third) genus within the family Adenoviridae. Arch Virol *143*, 829-837.

Bergelson, J.M., Cunningham, J.A., Droguett, G., Kurt-Jones, E.A., Krithivas, A., Hong, J.S., Horwitz, M.S., Crowell, R.L., and Finberg, R.W. (1997). Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. Science *275*, 1320-1323.

Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A *74*, 3171-3175.

Bergfors, T.M. (1999). Protein crystallization : techniques, strategies, and tips : a laboratory manual. (International University Line: La Jolla, California.).

Berk, A.J. (2007). Adenoviridae: The Viruses and Their replication. in Fields Virology. (Lippincott Williams & Wilkins: Philadelphia.).

Bewley, M.C., Springer, K., Zhang, Y.B., Freimuth, P., and Flanagan, J.M. (1999). Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. Science *286*, 1579-1583.

Buchholz, C.J., Koller, D., Devaux, P., Mumenthaler, C., Schneider-Schaulies, J., Braun, W., Gerlier, D., and Cattaneo, R. (1997). Mapping of the primary binding site of measles virus to its receptor CD46. J Biol Chem *272*, 22072-22079.

Burmeister, W.P., Guilligay, D., Cusack, S., Wadell, G., and Arnberg, N. (2004). Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. J Virol *78*, 7727-7736.

Carlson, H.C., al-Sheikhly, F., Pettit, J.R., and Seawright, G.L. (1974). Virus particles in spleens and intestines of turkeys with hemorrhagic enteritis. Avian Dis *18*, 67-73.

Chappell, J.D., Prota, A.E., Dermody, T.S., and Stehle, T. (2002). Crystal structure of reovirus attachment protein sigma1 reveals evolutionary relationship to adenovirus fiber. EMBO J *21*, 1-11.

Chatterjee, P.K., Vayda, M.E., and Flint, S.J. (1986). Identification of proteins and protein domains that contact DNA within adenovirus nucleoprotein cores by ultraviolet light crosslinking of oligonucleotides 32P-labelled in vivo. J Mol Biol *188*, 23-37.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr *66*, 12-21.

Chen, Y., Maguire, T., Hileman, R.E., Fromm, J.R., Esko, J.D., Linhardt, R.J., and Marks, R.M. (1997). Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. Nat Med *3*, 866-871.

Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell *12*, 1-8.

Chretien, I., Marcuz, A., Courtet, M., Katevuo, K., Vainio, O., Heath, J.K., White, S.J., and Du Pasquier, L. (1998). CTX, a Xenopus thymocyte receptor, defines a molecular family conserved throughout vertebrates. Eur J Immunol *28*, 4094-4104.

Chroboczek J., Ruigrok, R.W., Cusack S. (1995). Adenovirus fiber. Curr Top Microbiol Immunol *199*, 163-200.

Chroboczek, J., Viard, F., and D'Halluin, J.C. (1986). Human adenovirus 2 temperature-sensitive mutant 112 contains three mutations in the protein IIIa gene. Gene *49*, 157-160.

Cohen, C.J., Shieh, J.T., Pickles, R.J., Okegawa, T., Hsieh, J.T., and Bergelson, J.M. (2001). The coxsackievirus and adenovirus receptor is a transmembrane component of the tight junction. Proc Natl Acad Sci U S A *98*, 15191-15196.

Cohen, C.J., Xiang, Z.Q., Gao, G.P., Ertl, H.C., Wilson, J.M., and Bergelson, J.M. (2002). Chimpanzee adenovirus CV-68 adapted as a gene delivery vector interacts with the coxsackievirus and adenovirus receptor. J Gen Virol *83*, 151-155.

Colby, W.W., and Shenk, T. (1981). Adenovirus type 5 virions can be assembled in vivo in the absence of detectable polypeptide IX. J Virol *39*, 977-980.

Cupelli, K., and Stehle, T. (2011). Viral attachment strategies: the many faces of adenoviruses. Curr Opin Virol *1*, 84-91.

Davison, A.J., Wright, K.M., and Harrach, B. (2000). DNA sequence of frog adenovirus. J Gen Virol *81*, 2431-2439.

de Jong, J.C., Wigand, R., Kidd, A.H., Wadell, G., Kapsenberg, J.G., Muzerie, C.J., Wermenbol, A.G., and Firtzlaff, R.G. (1983). Candidate adenoviruses 40 and 41: fastidious adenoviruses from human infant stool. J Med Virol *11*, 215-231.

Dechecchi, M.C., Melotti, P., Bonizzato, A., Santacatterina, M., Chilosi, M., and Cabrini, G. (2001). Heparan sulfate glycosaminoglycans are receptors sufficient to mediate the initial binding of adenovirus types 2 and 5. J Virol *75*, 8772-8780.

Dechecchi, M.C., Tamanini, A., Bonizzato, A., and Cabrini, G. (2000). Heparan sulfate glycosaminoglycans are involved in adenovirus type 5 and 2-host cell interactions. Virology *268*, 382-390.

Di Guilmi, A.M., Barge, A., Kitts, P., Gout, E., and Chroboczek, J. (1995). Human adenovirus serotype 3 (Ad3) and the Ad3 fiber protein bind to a 130-kDa membrane protein on HeLa cells. Virus Res *38*, 71-81.

Domermuth, C.H., Gross, W.B., Douglass, C.S., DuBose, R.T., Harris, J.R., and Davis, R.B. (1977). Vaccination for hemorrhagic enteritis of turkeys. Avian Dis *21*, 557-565.

Domermuth, C.H., Harris, J.R., Gross, W.B., and Dubose, R.T. (1979). A naturally occurring infection of chickens with a hemorrhagic enteritis/marble spleen disease type of virus. . Avian Dis *23*, 479-484.

Dong, A., Xu, X., Edwards, A.M., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E.*, et al.* (2007). In situ proteolysis for protein crystallization and structure determination. Nat Methods *4*, 1019-1021.

Dormitzer, P.R., Sun, Z.Y., Blixt, O., Paulson, J.C., Wagner, G., and Harrison, S.C. (2002a). Specificity and affinity of sialic acid binding by the rhesus rotavirus VP8* core. J Virol *76*, 10512-10517.

Dormitzer, P.R., Sun, Z.Y., Wagner, G., and Harrison, S.C. (2002b). The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. EMBO J *21*, 885-897.

Echavarría, M. (2004). Adenoviruses. in Principles and Practice of Clinical Virology. (John Wiley & Sons Ltd: Chichester, West Sussex, England).

El Bakkouri, M., Seiradake, E., Cusack, S., Ruigrok, R.W., and Schoehn, G. (2008). Structure of the C-terminal head domain of the fowl adenovirus type 1 short fibre. Virology *378*, 169-176.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Enders, J.F., Bell, J.A., Dingle, J.H., Francis, T., Hilleman, M.R., Huebner, R.J., Payne, A.M. (1956). Adenoviruses: Group name proposed for new respiratory-tract viruses. Science *124*, 119-120.

Esko, J.D., Kimata, K., and Lindahl, U. (2009). Proteoglycans and Sulfated Glycosaminoglycans. In Essentials of Glycobiology. (Cold Spring Harbor Laboratory Press: New York, USA).

Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr D Biol Crystallogr *67*, 282-292.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr *69*, 1204-1214.

Everitt, E., Lutter, L., and Philipson, L. (1975). Structural proteins of adenoviruses. XII. Location and neighbor relationship among proteins of adenovirion type 2 as revealed by enzymatic iodination, immunoprecipitation and chemical cross-linking. Virology *67*, 197-208.

Everitt, E., and Philipson, L. (1974). Structural proteins of adenoviruses. XI. Purification of three low molecular weight virion proteins of adenovirus type 2 and their synthesis during productive infection. Virology *62*, 253-269.

Farkas, S.L., Harrach, B., and Benko, M. (2008). Completion of the genome analysis of snake adenovirus type 1, a representative of the reptilian lineage within the novel genus Atadenovirus. Virus Res *132*, 132-139.

Fender, P., Ruigrok, R.W., Gout, E., Buffet, S., and Chroboczek, J. (1997). Adenovirus dodecahedron, a new vector for human gene transfer. Nat Biotechnol *15*, 52-56.

Flint, S.J., Gallimore, P.H., and Sharp, P.A. (1975). Comparison of viral RNA sequences in adenovirus 2-transformed and lytically infected cells. J Mol Biol *96*, 47-68.

Flomenberg, P.R., Chen, M., Munk, G., and Horwitz, M.S. (1987). Molecular epidemiology of adenovirus type 35 infections in immunocompromised hosts. J Infect Dis *155*, 1127-1134.

Gaggar, A., Shayakhmetov, D.M., and Lieber, A. (2003). CD46 is a cellular receptor for group B adenoviruses. Nat Med *9*, 1408-1412.

Gaggar, A., Shayakhmetov, D.M., Liszewski, M.K., Atkinson, J.P., and Lieber, A. (2005). Localization of regions in CD46 that interact with adenovirus. J Virol *79*, 7503-7513.

Garon, C.F., Berry, K.W., Hierholzer, J.C., and Rose, J.A. (1973). Mapping of base sequence heterologies between genomes from different adenovirus serotypes. Virology *54*, 414-426.

Ghosh-Choudhury, G., Haj-Ahmad, Y., and Graham, F.L. (1987). Protein IX, a minor component of the human adenovirus capsid, is essential for the packaging of full length genomes. EMBO J *6*, 1733-1739.

Giannakis, E., Jokiranta, T.S., Ormsby, R.J., Duthy, T.G., Male, D.A., Christiansen, D., Fischetti, V.A., Bagley, C., Loveland, B.E., and Gordon, D.L. (2002). Identification of the streptococcal M protein binding site on membrane cofactor protein (CD46). J Immunol *168*, 4585-4592.

Gorman, J.J., Wallis, T.P., Whelan, D.A., Shaw, J., and Both, G.W. (2005). LH3, a "homologue" of the mastadenoviral E1B 55-kDa protein is a structural protein of atadenoviruses. Virology *342*, 159-166.

Greber, U.F., Willetts, M., Webster, P., and Helenius, A. (1993). Stepwise dismantling of adenovirus 2 during entry into cells. Cell *75*, 477-486.

Green, M., Mackey, J.K., Wold, W.S., and Rigden, P. (1979). Thirty-one human adenovirus serotypes (Ad1-Ad31) form five groups (A-E) based upon DNA genome homologies. Virology *93*, 481-492.

Green, N.M., Wrigley, N.G., Russell, W.C., Martin, S.R., and McLachlan, A.D. (1983). Evidence for a repeating cross-beta sheet structure in the adenovirus fibre. EMBO J *2*, 1357-1365.

Greenstone, H.L., Santoro, F., Lusso, P., and Berger, E.A. (2002). Human Herpesvirus 6 and Measles Virus Employ Distinct CD46 Domains for Receptor Function. J Biol Chem *277*, 39112-39118.

Guardado Calvo, P., Fox, G.C., Hermo Parrado, X.L., Llamas-Saiz, A.L., Costas, C., Martinez-Costas, J., Benavente, J., and van Raaij, M.J. (2005). Structure of the carboxy-terminal receptor-binding domain of avian reovirus fibre sigmaC. J Mol Biol *354*, 137-149.

Guardado-Calvo, P., Llamas-Saiz, A.L., Fox, G.C., Langlois, P., and van Raaij, M.J. (2007). Structure of the C-terminal head domain of the fowl adenovirus type 1 long fiber. J Gen Virol *88*, 2407-2416.

Guardado-Calvo, P., Munoz, E.M., Llamas-Saiz, A.L., Fox, G.C., Kahn, R., Curiel, D.T., Glasgow, J.N., and van Raaij, M.J. (2010). Crystallographic structure of porcine adenovirus type 4 fiber head and galectin domains. J Virol *84*, 10558-10568.

Guida, J.D., Fejer, G., Pirofski, L.A., Brosnan, C.F., and Horwitz, M.S. (1995). Mouse adenovirus type 1 causes a fatal hemorrhagic encephalomyelitis in adult C57BL/6 but not BALB/c mice. J Virol *69*, 7674-7681.

Gustin, K.E., Lutz, P., and Imperiale, M.J. (1996). Interaction of the adenovirus L1 52/55-kilodalton protein with the IVa2 gene product during infection. J Virol *70*, 6463-6467.

Hamelin, C., Jacques, C., and Lussier, G. (1988). Genome typing of mouse adenoviruses. J Clin Microbiol *26*, 31-33.

Hara, T., Kojima, A., Fukuda, H., Masaoka, T., Fukumori, Y., Matsumoto, M., and Seya, T. (1992). Levels of complement regulatory proteins, CD35 (CR1), CD46 (MCP) and CD55 (DAF) in human haematological malignancies. Br J Haematol *82*, 368-373.

Harrach, B., Benko, M., Both, G. W., Brown, M., Davison, A. J., Echavarrı´a, M., Hess, M., Jones, M. S., Kajon, A., Lehmkuhl, H. D., Mautner, V., Mittal, S. K. &Wadell, G. (2011). Virus Taxonomy: Classification and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses, edited by A M Q King, M J Adams, E B Carstens & E JLefkowitz, pp 125–141 (Elsevier: San Diego, USA).

Hartley, J.W., and Rowe, W.P. (1960). A new mouse virus apparently related to the adenovirus group. Virology *11*, 645-647.

Hashimoto, K., Sugiyama, T., and Sasaki, S. (1966). An adenovirus isolated from the feces of mice I. Isolation and identification. Jpn J Microbiol *10*, 115-125.

Hemmi, S., Vidovszky, M.Z., Ruminska, J., Ramelli, S., Decurtins, W., Greber, U.F., and Harrach, B. (2011). Genomic and phylogenetic analyses of murine adenovirus 2. Virus Res *160*, 128-135.

Hendrickson, W.A. (1985). Analysis of Protein Structure from Diffraction Measurement at Multiple Wavelengths. Trans ACA *21*.

Henry, L.J., Xia, D., Wilke, M.E., Deisenhofer, J., and Gerard, R.D. (1994). Characterization of the knob domain of the adenovirus type 5 fiber protein expressed in Escherichia coli. J Virol *68*, 5239-5246.

Hess, M., Cuzange, A., Ruigrok, R.W., Chroboczek, J., and Jacrot, B. (1995). The avian adenovirus penton: two fibres and one base. J Mol Biol *252*, 379-385.

Hidaka, C., Milano, E., Leopold, P.L., Bergelson, J.M., Hackett, N.R., Finberg, R.W., Wickham, T.J., Kovesdi, I., Roelvink, P., and Crystal, R.G. (1999). CAR-dependent and CAR-independent pathways of adenovirus vector-mediated gene transfer and expression in human fibroblasts. J Clin Invest *103*, 579-587.

Hierholzer, J.C. (1973). Further subgrouping of the human adenoviruses by differential hemagglutination. J Infect Dis *128*, 541-550.

Hilleman, M.R., and Werner, J.H. (1954). Recovery of new agent from patients with acute respiratory illness. Proc Soc Exp Biol Med *85*, 183-188.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res *38*, W545-549.

Holyoak, T., Fenn, T.D., Wilson, M.A., Moulin, A.G., Ringe, D., and Petsko, G.A. (2003). Malonate: a versatile cryoprotectant and stabilizing solution for salt-grown macromolecular crystals. Acta Crystallogr D Biol Crystallogr *59*, 2356-2358.

Hosokawa, K., and Sung, M.T. (1976). Isolation and characterization of an extremely basic protein from adenovirus type 5. J Virol *17*, 924-934.

Howell, P.L., and Smith, G.D. (1992). Identification of heavy-atom derivatives by normal probability methods. J Appl Cryst *25*, 81–86.

Hsu, E.C., Dorig, R.E., Sarangi, F., Marcil, A., Iorio, C., and Richardson, C.D. (1997). Artificial mutations and natural variations in the CD46 molecules from human and monkey cells define regions important for measles virus binding. J Virol *71*, 6144-6154.

Huebner, R.J., Rowe, W.P., Ward, T.G., Parrott, R.H., and Bell, J.A. (1954). Adenoidal-pharyngeal-conjunctival agents: a newly recognized group of common viruses of the respiratory system. N Engl J Med *251*, 1077-1086.

Itakura, C., and Carlson, H.C. (1975). Electron microscopic findings of cells with inclusion bodies in experimental hemorrhagic enteritis of turkeys. Can J Comp Med *39*, 299-304.

Jawetz, E., Kimura, S.J., Hanna, L., Coleman, V.R., Thygeson, P., and Nicholas, A. (1955). Studies on the etiology of epidemic keratoconjunctivitis. Am J Ophthalmol *40*, 200-209; discussion, 209-211.

Jawetz, E., Thygeson, P., Hanna, L., Nicholas, A., and Kimura, S.J. (1957). The etiology of epidemic keratoconjunctivitis. Am J Ophthalmol *43*, 79-83.

Juhasz, A., and Ahne, W. (1993). Physicochemical properties and cytopathogenicity of an adenovirus-like agent isolated from corn snake (Elaphe guttata). Arch Virol *130*, 429-439.

Julian, A.F., and Durham, P.J. (1982). Adenoviral hepatitis in a female bearded dragon (Amphibolurus barbatus). N Z Vet J *30*, 59-60.

Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.

Kajon, A.E., and Wadell, G. (1996). Sequence analysis of the E3 region and fiber gene of human adenovirus genome type 7h. Virology *215*, 190-196.

Kalyuzhniy, O., Di Paolo, N.C., Silvestry, M., Hofherr, S.E., Barry, M.A., Stewart, P.L., and Shayakhmetov, D.M. (2008). Adenovirus serotype 5 hexon is critical for virus infection of hepatocytes in vivo. Proc Natl Acad Sci U S A *105*, 5483-5488.

Kidd, A.H., Chroboczek, J., Cusack, S., and Ruigrok, R.W. (1993). Adenovirus type 40 virions contain two distinct fibers. Virology *192*, 73-84.

Kilcoyne, M., Gerlach, J.Q., Kane, M., and Joshi, L. (2012). Surface chemistry and linker effects on lectin-carbohydrate recognition for glycan microarrays. Anal Methods *4*, 2721-2728.

Kinchington, P.R., Romanowski, E.G., and Jerold Gordon, Y. (2005). Prospects for adenovirus antivirals. J Antimicrob Chemother *55*, 424-429.

Kirchner, E., Guglielmi, K.M., Strauss, H.M., Dermody, T.S., and Stehle, T. (2008). Structure of reovirus sigma1 in complex with its receptor junctional adhesion molecule-A. PLoS Pathog *4*, e1000235.

Kleiboeker, S.B. (1995). Sequence analysis of the fiber genomic region of a porcine adenovirus predicts a novel fiber protein. Virus Res *39*, 299-309.

Klein, M., Earley, E., Zellat, J. (1959). Isolation from cattle of a virus related to human adenovirus. Proc Soc Exp Biol Med *102*, 1–4.

Klempa, B., Kruger, D.H., Auste, B., Stanko, M., Krawczyk, A., Nickel, K.F., Uberla, K., and Stang, A. (2009). A novel cardiotropic murine adenovirus representing a distinct species of mastadenoviruses. J Virol *83*, 5749-5759.

Kobayashi, M., Kubota, M., and Matsuura, Y. (1999). Crystallization and improvement of crystal quality for x-ray diffraction of maltooligosyl trehalose synthase by reductive methylation of lysine residues. Acta Crystallogr D Biol Crystallogr *55*, 931-933.

Kovacs, E.R., and Benko, M. (2011). Complete sequence of raptor adenovirus 1 confirms the characteristic genome organization of siadenoviruses. Infect Genet Evol *11*, 1058-1065.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J Mol Biol *372*, 774-797.

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc *3*, 1171-1179.

Lavinder, J.J., Hari, S.B., Sullivan, B.J., and Magliery, T.J. (2009). High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. J Am Chem Soc *131*, 3794-3795.

Legrand, V., Spehner, D., Schlesinger, Y., Settelen, N., Pavirani, A., and Mehtali, M. (1999). Fiberless recombinant adenoviruses: virus maturation and infectivity in the absence of fiber. J Virol *73*, 907-919.

Lenaerts, L., McVey, J.H., Baker, A.H., Denby, L., Nicklin, S., Verbeken, E., and Naesens, L. (2009). Mouse adenovirus type 1 and human adenovirus type 5 differ in endothelial cell tropism and liver targeting. J Gene Med *11*, 119-127.

Leopold, P.L., and Crystal, R.G. (2007). Intracellular trafficking of adenovirus: many means to many ends. Adv Drug Deliv Rev *59*, 810-821.

Li, E., Brown, S.L., Stupack, D.G., Puente, X.S., Cheresh, D.A., and Nemerow, G.R. (2001). Integrin alpha(v)beta1 is an adenovirus coreceptor. J Virol *75*, 5405-5409.

Liszewski, M.K., Farries, T.C., Lublin, D.M., Rooney, I.A., and Atkinson, J.P. (1996). Control of the complement system. Adv Immunol *61*, 201-283.

Liszewski, M.K., Leung, M., Cui, W., Subramanian, V.B., Parkinson, J., Barlow, P.N., Manchester, M., and Atkinson, J.P. (2000). Dissecting sites important for complement regulatory activity in membrane cofactor protein (MCP; CD46). J Biol Chem *275*, 37692-37701.

Littlechild, J.A. (1991). Protein crystallization: magical or logical: can we establish some general rules? J Phys D: Appl Phys *24*.

Liu, G.Q., Babiss, L.E., Volkert, F.C., Young, C.S., and Ginsberg, H.S. (1985). A thermolabile mutant of adenovirus 5 resulting from a substitution mutation in the protein VIII gene. J Virol *53*, 920-925.

Liu, H., Jin, L., Koh, S.B., Atanasov, I., Schein, S., Wu, L., and Zhou, Z.H. (2010). Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. Science *329*, 1038-1043.

Lortat-Jacob, H., Chouin, E., Cusack, S., and van Raaij, M.J. (2001). Kinetic analysis of adenovirus fiber binding to its receptor reveals an avidity mechanism for trimeric receptor-ligand interactions. J Biol Chem *276*, 9009-9015.

Löser, P., G. Cichon, G. Jennings, G. Both, and Hofmann, C. (1999). Ovine adenovirus vectors promote efficient gene delivery in vivo. Gene Ther Mol Biol *4*, 33-43.

Loser, P., Kumin, D., Hillgenberg, M., Both, G.W., and Hofmann, C. (2002). Preparation of ovine adenovirus vectors. Methods Mol Med *69*, 415-426.

Louis, N., Fender, P., Barge, A., Kitts, P., and Chroboczek, J. (1994). Cell-binding domain of adenovirus serotype 2 fiber. J Virol *68*, 4104-4106.

Maier, O., Galan, D.L., Wodrich, H., and Wiethoff, C.M. (2010). An N-terminal domain of adenovirus protein VI fragments membranes by inducing positive membrane curvature. Virology *402*, 11-19.

Maizel, J.V., Jr., White, D.O., and Scharff, M.D. (1968). The polypeptides of adenovirus. I. Evidence for multiple protein components in the virion and a comparison of types 2, 7A, and 12. Virology *36*, 115-125.

Marschang, R.E. (2011). Viruses infecting reptiles. Viruses *3*, 2087-2126.

Martin, M.A., Knipe, D.M., Fields, B.N., Howley, P.M., Griffin, D., and Lamb, R. (2007). Fields' virology. (Wolters Kluwer Health/Lippincott Williams & Wilkins: Philadelphia.).

Matrosovich, M., Herrler, G., and Klenk, H.D. (2013). Sialic Acid Receptors of Viruses. Top Curr Chem.

Matthews, B.W. (1968). Solvent content of protein crystals. J Mol Biol *33*, 491-497.

Matthews, B.W. (1985). Determination of protein molecular weight, hydration, and packing from crystal density. Methods Enzymol *114*, 176-187.

Matthews, D.A., and Russell, W.C. (1998). Adenovirus core protein V is delivered by the invading virus to the nucleus of the infected cell and later in infection is associated with nucleoli. J Gen Virol *79 (Pt 7)*, 1671-1675.

Mautner, V., and Pereira, H.G. (1971). Crystallization of a second adenovirus protein (the fibre). Nature *230*, 456-457.

Mautner, V., Steinthorsdottir, V., and Bailey, A. (1995). Enteric adenoviruses. Curr Top Microbiol Immunol *199 (Pt 3)*, 229-282.

Mayer, M., and Meyer, B. (1999). Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. Angew Chem Int Ed *38*, 1784–1788.

McConnell, M.J., and Imperiale, M.J. (2004). Biology of adenovirus and its use as a vector for gene therapy. Hum Gene Ther *15*, 1022-1033.

McCoy, A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr D Biol Crystallogr *63*, 32-41.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404-405.

McNicholas, S., Potterton, E., Wilson, K.S., and Noble, M.E. (2011). Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr D Biol Crystallogr *67*, 386-394.

McPherson, A. (2008). Introduction to Macromolecular Crystallography (John Wiley & Sons, Inc: New Jersey, USA).

Mei, Y.F., Lindman, K., and Wadell, G. (2002). Human adenoviruses of subgenera B, C, and E with various tropisms differ in both binding to and replication in the epithelial A549 and 293 cells. Virology *295*, 30-43.

Meier, O., and Greber, U.F. (2003). Adenovirus endocytosis. J Gene Med *5*, 451-462.

Meissner, J.D., Hirsch, G.N., LaRue, E.A., Fulcher, R.A., and Spindler, K.R. (1997). Completion of the DNA sequence of mouse adenovirus type 1: sequence of E2B, L1, and L2 (18-51 map units). Virus Res *51*, 53-64.

Menendez-Conejero, R. (2013). PhD thesis. (Universidad Autonoma de Madrid, Spain).

Merckel, M.C., Huiskonen, J.T., Bamford, D.H., Goldman, A., and Tuma, R. (2005). The structure of the bacteriophage PRD1 spike sheds light on the evolution of viral capsid architecture. Mol Cell *18*, 161-170.

Meyer, B., and Peters, T. (2003). NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angew Chem Int Ed *42*, 864-890.

Mitraki, A., Barge, A., Chroboczek, J., Andrieu, J.P., Gagnon, J., and Ruigrok, R.W. (1999). Unfolding studies of human adenovirus type 2 fibre trimers. Evidence for a stable domain. Eur J Biochem *264*, 599-606.

Mitraki, A., Papanikolopoulou, K., and Van Raaij, M.J. (2006). Natural triple beta-stranded fibrous folds. Adv Protein Chem *73*, 97-124.

Murray, K.P., Mathure, S., Kaul, R., Khan, S., Carson, L.F., Twiggs, L.B., Martens, M.G., and Kaul, A. (2000). Expression of complement regulatory proteins-CD 35, CD 46, CD 55, and CD 59-in benign and malignant endometrial tissue. Gynecol Oncol *76*, 176-182.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr *67*, 355-367.

Nemerow, G.R., Pache, L., Reddy, V., and Stewart, P.L. (2009). Insights into adenovirus host cell interactions from structural studies. Virology *384*, 380-388.

Nicklin, S.A., Wu, E., Nemerow, G.R., and Baker, A.H. (2005). The influence of adenovirus fiber structure and function on vector development for gene therapy. Mol Ther *12*, 384-393.

Nilsson, E.C., Storm, R.J., Bauer, J., Johansson, S.M., Lookene, A., Angstrom, J., Hedenstrom, M., Eriksson, T.L., Frangsmyr, L., Rinaldi, S.*, et al.* (2011). The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. Nat Med *17*, 105-109.

Noureddini, S.C., and Curiel, D.T. (2005). Genetic targeting strategies for adenovirus. Mol Pharm *2*, 341-347.

Ogawa, M., Ahne, W., and Essbauer, S. (1992). Reptilian viruses: adenovirus-like agent isolated from royal python (Python regius). Zentralbl Veterinarmed B *39*, 732-736.

Palya, V., Nagy, M., Glavits, R., Ivanics, E., Szalay, D., Dan, A., Suveges, T., Markos, B., and Harrach, B. (2007). Investigation of field outbreaks of turkey haemorrhagic enteritis in Hungary. Acta Vet Hung *55*, 135-149.

Pantelic, R.S., Lockett, L.J., Rothnagel, R., Hankamer, B., and Both, G.W. (2008). Cryoelectron microscopy map of Atadenovirus reveals cross-genus structural differences from human adenovirus. J Virol *82*, 7346-7356.

Papp, T., Fledelius, B., Schmidt, V., Kajan, G.L., and Marschang, R.E. (2009). PCR-sequence characterization of new adenoviruses found in reptiles and the first successful isolation of a lizard adenovirus. Vet Microbiol *134*, 233-240.

Park, Y.M., Kim, J.H., Gu, S.H., Lee, S.Y., Lee, M.G., Kang, Y.K., Kang, S.H., Kim, H.J., and Song, J.W. (2012). Full genome analysis of a novel adenovirus from the South Polar skua (Catharacta maccormicki) in Antarctica. Virology *422*, 144-150.

Penzes, J.J., Menendez-Conejero, R., Condezo, G.N., Ball, I., Papp, T., Doszpoly, A., Paradela, A., Perez-Berna, A.J., Lopez-Sanz, M., Nguyen, T.H.*, et al.* (2014). Molecular Characterization of a Lizard Adenovirus Reveals The First Atadenovirus with two Fiber Genes, and the First Adenovirus with Either One Short or Three Long Fibers per Penton. J Virol *88*, 11304-11314.

Pereira, H.G., Valentine, R.C., and Russell, W.C. (1968). Crystallization of an adenovirus protein (the hexon). Nature *219*, 946-947.

Persson, B.D., Reiter, D.M., Marttila, M., Mei, Y.F., Casasnovas, J.M., Arnberg, N., and Stehle, T. (2007). Adenovirus type 11 binding alters the conformation of its receptor CD46. Nat Struct Mol Biol *14*, 164-166.

Persson, B.D., Schmitz, N.B., Santiago, C., Zocher, G., Larvie, M., Scheu, U., Casasnovas, J.M., and Stehle, T. (2010). Structure of the extracellular portion of CD46 provides insights into its interactions with complement proteins and pathogens. PLoS Pathog *6*, e1001122.

Philipson, L. (1995). Adenovirus--an eternal archetype. Curr Top Microbiol Immunol *199 (Pt 1)*, 1-24.

Raman, S., Hsu, T.H., Ashley, S.L., and Spindler, K.R. (2009). Usage of integrin and heparan sulfate as receptors for mouse adenovirus type 1. J Virol *83*, 2831-2838.

Reiter, D.M., Frierson, J.M., Halvorson, E.E., Kobayashi, T., Dermody, T.S., and Stehle, T. (2011). Crystal structure of reovirus attachment protein sigma1 in complex with sialylated oligosaccharides. PLoS Pathog *7*, e1002166.

Rekosh, D.M., Russell, W.C., Bellet, A.J., and Robinson, A.J. (1977). Identification of a protein linked to the ends of adenovirus DNA. Cell *11*, 283-295.

Roberts, E., Eargle, J., Wright, D., and Luthey-Schulten, Z. (2006). MultiSeq: unifying sequence and structure data for evolutionary analysis. BMC Bioinformatics *7*, 382.

Roelvink, P.W., Mi Lee, G., Einfeld, D.A., Kovesdi, I., and Wickham, T.J. (1999). Identification of a conserved receptor-binding site on the fiber proteins of CAR-recognizing adenoviridae. Science *286*, 1568-1571.

Rosen, L. (1958). Hemagglutination by adenoviruses. Virology *5*, 574-577.

Rossmann, M.G.B., D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. Acta Crystallogr *15*, 24-31.

Rostand, K.S., and Esko, J.D. (1997). Microbial adherence to and invasion through proteoglycans. Infect Immun *65*, 1-8.

Rowe, W.P., Huebner, R.J., Gilmore, L.K., Parrott, R.H., and Ward, T.G. (1953). Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. Proc Soc Exp Biol Med *84*, 570-573.

Russell, W.C. (2000). Update on adenovirus and its vectors. J Gen Virol *81*, 2573-2604.

Russell, W.C., Laver, W.G., and Sanderson, P.J. (1968). Internal components of adenovirus. Nature *219*, 1127-1130.

Russell, W.C., and Precious, B. (1982). Nucleic acid-binding properties of adenovirus structural polypeptides. J Gen Virol *63 (Pt 1)*, 69-79.

Rux, J.J., and Burnett, R.M. (2004). Adenovirus structure. Hum Gene Ther *15*, 1167-1176.

San Martin, C. (2012). Latest insights on adenovirus structure and assembly. Viruses *4*, 847-877.

San Martin, C., Glasgow, J.N., Borovjagin, A., Beatty, M.S., Kashentseva, E.A., Curiel, D.T., Marabini, R., and Dmitriev, I.P. (2008). Localization of the N-terminus of minor coat protein IIIa in the adenovirus capsid. J Mol Biol *383*, 923-934.

Sauter, N.K., Bednarski, M.D., Wurzburg, B.A., Hanson, J.E., Whitesides, G.M., Skehel, J.J., and Wiley, D.C. (1989). Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: a 500-MHz proton nuclear magnetic resonance study. Biochemistry *28*, 8388-8396.

Segerman, A., Atkinson, J.P., Marttila, M., Dennerquist, V., Wadell, G., and Arnberg, N. (2003). Adenovirus type 11 uses CD46 as a cellular receptor. J Virol *77*, 9183-9191.

Seiradake, E., Henaff, D., Wodrich, H., Billet, O., Perreau, M., Hippert, C., Mennechet, F., Schoehn, G., Lortat-Jacob, H., Dreja, H*., et al.* (2009). The cell adhesion molecule "CAR" and sialic acid on human erythrocytes influence adenovirus in vivo biodistribution. PLoS Pathog *5*, e1000277.

Seiradake, E., Lortat-Jacob, H., Billet, O., Kremer, E.J., and Cusack, S. (2006). Structural and mutational analysis of human Ad37 and canine adenovirus 2 fiber heads in complex with the D1 domain of coxsackie and adenovirus receptor. J Biol Chem *281*, 33704-33716.

Shayakhmetov, D.M., and Lieber, A. (2000). Dependence of adenovirus infectivity on length of the fiber shaft domain. J Virol *74*, 10274-10286.

Sheldrick, G.M. (2008). A short history of SHELX. Acta Crystallogr A *64*, 112-122.

Shenk, T. (1996). Adenoviridae: The viruses and their replication. (Lippincott-Raven Publishers: Philadelphia, PA, USA).

Short, J.J., Vasu, C., Holterman, M.J., Curiel, D.T., and Pereboev, A. (2006). Members of adenovirus species B utilize CD80 and CD86 as cellular attachment receptors. Virus Res *122*, 144-153.

Signas, C., Akusjarvi, G., and Pettersson, U. (1985). Adenovirus 3 fiber polypeptide gene: implications for the structure of the fiber protein. J Virol *53*, 672-678.

Silim, A., Thorsen, J., and Carlson, H.C. (1978). Experimental infection of chickens with hemorrhagic enteritis virus. Avian Dis *22*, 106-114.

Singh, A.K., Ballmann, M.Z., Benko, M., Harrach, B., and van Raaij, M.J. (2013a). Crystallization of the C-terminal head domain of the fibre protein from a siadenovirus, turkey adenovirus 3. Acta Crystallogr Sect F Struct Biol Cryst Commun *69*, 1135-1139.

Singh, A.K., Menendez-Conejero, R., San Martin, C., and van Raaij, M.J. (2013b). Crystallization of the C-terminal domain of the fibre protein from snake adenovirus 1, an atadenovirus. Acta Crystallogr Sect F Struct Biol Cryst Commun *69*, 1374-1379.

Singh, M., Shmulevitz, M., and Tikoo, S.K. (2005). A newly identified interaction between IVa2 and pVIII proteins during porcine adenovirus type 3 infection. Virology *336*, 60-69.

Sirena, D., Lilienfeld, B., Eisenhut, M., Kalin, S., Boucke, K., Beerli, R.R., Vogt, L., Ruedl, C., Bachmann, M.F., Greber, U.F.*, et al.* (2004). The human membrane cofactor CD46 is a receptor for species B adenovirus serotype 3. J Virol *78*, 4454-4462.

Smith, T.A., Idamakanti, N., Rollence, M.L., Marshall-Neff, J., Kim, J., Mulgrew, K., Nemerow, G.R., Kaleko, M., and Stevenson, S.C. (2003). Adenovirus serotype 5 fiber shaft influences in vivo gene transfer in mice. Hum Gene Ther *14*, 777-787.

Sohier, R., Chardonnet, Y., and Prunieras, M. (1965). Adenoviruses. Status of current knowledge. Prog Med Virol *7*, 253-325.

Sorensen, E.B., and Mesner, P.W., Jr. (2005). IgH-2 cells: a reptilian model for apoptotic studies. Comp Biochem Physiol B Biochem Mol Biol *140*, 163-170.

Soudais, C., Boutin, S., Hong, S.S., Chillon, M., Danos, O., Bergelson, J.M., Boulanger, P., and Kremer, E.J. (2000). Canine adenovirus type 2 attachment and internalization: coxsackievirus-adenovirus receptor, alternative receptors, and an RGD-independent pathway. J Virol *74*, 10639-10649.

Spinelli, S., Desmyter, A., Verrips, C.T., de Haard, H.J., Moineau, S., and Cambillau, C. (2006). Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. Nat Struct Mol Biol *13*, 85-89.

Spjut, S., Qian, W., Bauer, J., Storm, R., Frangsmyr, L., Stehle, T., Arnberg, N., and Elofsson, M. (2011). A potent trivalent sialic acid inhibitor of adenovirus type 37 infection of human corneal cells. Angew Chem Int Ed Engl *50*, 6519-6521.

Stehle, T., and Harrison, S.C. (1997). High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. EMBO J *16*, 5139-5148.

Stewart, P.L., and Burnett, R.M. (1993). Adenovirus Structure as Revealed by X-Ray Crystallography, Electron Microscopy, and Difference Imaging. Jpn J Appl Phys *32*, 1342.

Stewart, P.L., Fuller, S.D., and Burnett, R.M. (1993). Difference imaging of adenovirus: bridging the resolution gap between X-ray crystallography and electron microscopy. EMBO J *12*, 2589-2599.

Stewart, P.L., and Nemerow, G.R. (2007). Cell integrins: commonly used receptors for diverse viral pathogens. Trends Microbiol *15*, 500-507.

Sugiyama, T., Hashimoto, K., and Sasaki, S. (1967). An adenovirus isolated from the feces of mice. II. Experimental infection. Jpn J Microbiol *11*, 33-42.

Takeuchi, A., and Hashimoto, K. (1976). Electron microscope study of experimental enteric adenovirus infection in mice. Infect Immun *13*, 569-580.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol *30*, 2725-2729.

Tan, P.K., Michou, A.I., Bergelson, J.M., and Cotten, M. (2001). Defining CAR as a cellular receptor for the avian adenovirus CELO using a genetic analysis of the two viral fibre proteins. J Gen Virol *82*, 1465-1472.

Templeton, D.M. (1992). Proteoglycans in cell regulation. Crit Rev Clin Lab Sci *29*, 141-184.

Tolin, S.A., and Domermuth, C.H. (1975). Hemorrhagic enteritis of turkeys: electron microscopy of the causal virus. Avian Dis *19*, 118-125.

Tomko, R.P., Johansson, C.B., Totrov, M., Abagyan, R., Frisen, J., and Philipson, L. (2000). Expression of the adenovirus receptor and its interaction with the fiber knob. Exp Cell Res *255*, 47-55.

Trentin, J.J., Yabe, Y., and Taylor, G. (1962). The quest for human cancer viruses. Science *137*, 835-841.

Trotman, L.C., Mosberger, N., Fornerod, M., Stidwill, R.P., and Greber, U.F. (2001). Import of adenovirus DNA involves the nuclear pore complex receptor CAN/Nup214 and histone H1. Nat Cell Biol *3*, 1092-1100.

van den Hurk, J.V. (1992). Characterization of the structural proteins of hemorrhagic enteritis virus. Arch Virol *126*, 195-213.

van Oostrum, J., and Burnett, R.M. (1985). Molecular composition of the adenovirus type 2 virion. J Virol *56*, 439-448.

van Raaij, M.J., Chouin, E., van der Zandt, H., Bergelson, J.M., and Cusack, S. (2000). Dimeric structure of the coxsackievirus and adenovirus receptor D1 domain at 1.7 A resolution. Structure *8*, 1147-1155.

van Raaij, M.J., Louis, N., Chroboczek, J., and Cusack, S. (1999a). Structure of the human adenovirus serotype 2 fiber head domain at 1.5 A resolution. Virology *262*, 333-343.

van Raaij, M.J., Mitraki, A., Lavigne, G., and Cusack, S. (1999b). A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. Nature *401*, 935-938.

Veesler, D., Spinelli, S., Mahony, J., Lichiere, J., Blangy, S., Bricogne, G., Legrand, P., Ortiz-Lombardia, M., Campanacci, V., van Sinderen, D.*, et al.* (2012). Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism. Proc Natl Acad Sci U S A *109*, 8954-8958.

Vellinga, J., Van der Heijdt, S., and Hoeben, R.C. (2005). The adenovirus capsid: major progress in minor proteins. J Gen Virol *86*, 1581-1588.

Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods Mol Biol *364*, 215-230.

Vrati, S., Brookes, D.E., Strike, P., Khatri, A., Boyle, D.B., and Both, G.W. (1996). Unique genome arrangement of an ovine adenovirus: identification of new proteins and proteinase cleavage sites. Virology *220*, 186-199.

Waddington, S.N., McVey, J.H., Bhella, D., Parker, A.L., Barker, K., Atoda, H., Pink, R., Buckley, S.M., Greig, J.A., Denby, L.*, et al.* (2008). Adenovirus serotype 5 hexon mediates liver gene transfer. Cell *132*, 397-409.

Wadell, G. (1979). Classification of human adenoviruses by SDS-polyacrylamide gel electrophoresis of structural polypeptides. Intervirology *11*, 47-57.

Wadell, G., Hammarskjold, M.L., Winberg, G., Varsanyi, T.M., and Sundell, G. (1980). Genetic variability of adenoviruses. Ann N Y Acad Sci *354*, 16-42.

Walter, T.S., Meier, C., Assenberg, R., Au, K.F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., and Grimes, J.M. (2006). Lysine methylation as a routine rescue strategy for protein crystallization. Structure *14*, 1617-1622.

Walters, R.W., Freimuth, P., Moninger, T.O., Ganske, I., Zabner, J., and Welsh, M.J. (2002). Adenovirus fiber disrupts CAR-mediated intercellular adhesion allowing virus escape. Cell *110*, 789-799.

Wang, H., Li, Z.Y., Liu, Y., Persson, J., Beyer, I., Moller, T., Koyuncu, D., Drescher, M.R., Strauss, R., Zhang, X.B.*, et al.* (2011). Desmoglein 2 is a receptor for adenovirus serotypes 3, 7, 11 and 14. Nat Med *17*, 96-104.

Wang, H., Liaw, Y.C., Stone, D., Kalyuzhniy, O., Amiraslanov, I., Tuve, S., Verlinde, C.L., Shayakhmetov, D., Stehle, T., Roffler, S.*, et al.* (2007). Identification of CD46 binding sites within the adenovirus serotype 35 fiber knob. J Virol *81*, 12785-12792.

Wang, X., and Bergelson, J.M. (1999). Coxsackievirus and adenovirus receptor cytoplasmic and transmembrane domains are not essential for coxsackievirus and adenovirus infection. J Virol *73*, 2559-2562.

Weis, W., Brown, J.H., Cusack, S., Paulson, J.C., Skehel, J.J., and Wiley, D.C. (1988). Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. Nature *333*, 426-431.

Wellehan, J.F., Johnson, A.J., Harrach, B., Benko, M., Pessier, A.P., Johnson, C.M., Garner, M.M., Childress, A., and Jacobson, E.R. (2004). Detection and analysis of six lizard adenoviruses by consensus primer PCR provides further evidence of a reptilian origin for the atadenoviruses. J Virol *78*, 13366-13369.

Wickham, T.J., Mathias, P., Cheresh, D.A., and Nemerow, G.R. (1993). Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. Cell *73*, 309-319.

Wiethoff, C.M., Wodrich, H., Gerace, L., and Nemerow, G.R. (2005). Adenovirus protein VI mediates membrane disruption following capsid disassembly. J Virol *79*, 1992-2000.

Wigand, R. (1980). Age and susceptibility of Swiss mice for mouse adenovirus, strain FL. Arch Virol *64*, 349-357.

Wigand, R., Gelderblom, H., and Wadell, G. (1980). New human adenovirus (candidate adenovirus 36), a novel member of subgroup D. Arch Virol *64*, 225-233.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr *67*, 235-242.

Winter, G., Lobley, C.M., and Prince, S.M. (2013). Decision making in xia2. Acta Crystallogr D Biol Crystallogr *69*, 1260-1273.

Wodrich, H., Guan, T., Cingolani, G., Von Seggern, D., Nemerow, G., and Gerace, L. (2003). Switch from capsid protein import to adenovirus assembly by cleavage of nuclear transport signals. EMBO J *22*, 6245-6255.

Wu, E., Pache, L., Von Seggern, D.J., Mullen, T.M., Mikyas, Y., Stewart, P.L., and Nemerow, G.R. (2003). Flexibility of the adenovirus fiber is required for efficient receptor interaction. J Virol *77*, 7225-7235.

Wu, E., Trauger, S.A., Pache, L., Mullen, T.M., von Seggern, D.J., Siuzdak, G., and Nemerow, G.R. (2004a). Membrane cofactor protein is a receptor for adenoviruses associated with epidemic keratoconjunctivitis. J Virol *78*, 3897-3905.

Wu, Q., Chen, Y., Kulshreshtha, V., and Tikoo, S.K. (2004b). Characterization and nuclear localization of the fiber protein encoded by the late region 7 of bovine adenovirus type 3. Arch Virol *149*, 1783-1799.

Xia, D., Henry, L.J., Gerard, R.D., and Deisenhofer, J. (1994). Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1.7 A resolution. Structure *2*, 1259-1270.

Yabe, Y., Samper, L., Bryan, E., Taylor, G., and Trentin, J.J. (1964). Oncogenic Effect of Human Adenovirus Type 12, in Mice. Science *143*, 46-47.

Zhang, W., and Arcos, R. (2005). Interaction of the adenovirus major core protein precursor, pVII, with the viral DNA packaging machinery. Virology *334*, 194-202.

Zhang, Y., and Bergelson, J.M. (2005). Adenovirus receptors. J Virol *79*, 12125-12131.

Zubieta, C., Schoehn, G., Chroboczek, J., and Cusack, S. (2005). The structure of the human adenovirus 2 penton. Mol Cell *17*, 121-135.

Annex

8. Annex

8.1 List of publications

Singh AK, Menéndez-Conejero R, San Martín C, van Raaij MJ (2013) Crystallization of the C-terminal domain of the fibre protein from snake adenovirus 1, an atadenovirus. Acta Cryst. F**69**:1374-1379 doi: 10.1107/S1744309113029308.

Singh AK, Ballmann MZ, Benkő M, Harrach B, van Raaij MJ (2013) Crystallization of the C-terminal head domain of the fibre protein from a siadenovirus, turkey adenovirus 3. Acta Cryst. F**69**:1135-1139. doi:10.1107/S174430911302397X.

Singh AK, Berbís Á, Jiménez-Barbero J, Menéndez M, Kilcoyne M, Joshi L, Ballmann MZ, Benkő M, Harrach B, van Raaij MJ. Structure and sialyl-lactose binding of the carboxy-terminal head domain of the fibre from a Siadenovirus, turkey adenovirus 3 (*manuscript in preparation*).

Singh AK, Menéndez-Conejero R, San Martín C, van Raaij MJ. Crystal structure of the fibre head domain of the Atadenovirus snake adenovirus 1 (*submitted*).

Granell M, Namura M, Alvira S, Garcia-Doval C, **Singh AK**, Gutsche I, van Raaij MJ, Kanamaru S (2014). Crystallization of the carboxy-terminal region of the bacteriophage T4 proximal long tail fibre protein gp34. Acta Cryst. F**70**:970–975 doi: 10.1107/S2053230X14010449.

Nikhil K, Sharan S, **Singh AK**, Chakraborty A, Roy P (2014) Anticancer activities of pterostilbene-isothiocyanate conjugate in breast cancer cells: Involvement of PPARγ. PLoS ONE 9(8): e104592. doi:10.1371/journal.pone.0104592.

Table AI: List of glycoproteins and neoglycoconjugates used in glycan array screenings

Abbreviation	Neoglycoconjugate
Fetuin	Fetuin
ASF	Asialofetuin
PBS	Phosphate buffered saline
Ov	Ovalbumin
RB	RNAse B
Xferrin	Transferrin
4APHSA	4AP-HAS
a-C	a-Crystallin from bovine lens
M3BSA	Mana1,3(Mana1,6)Man-BSA
GlcNAcBSA	GlcNAc-BSA
LacNAcBSA	LacNAc-BSA
3SLNBSA	3'SialylLacNAc-BSA
3SLacHSA	3'-Sialyllactose-APD-HSA
6SLacHSA	6'-Sialyllactose-APD-HSA
2FLBSA	2'Fucosyllactose-BSA
3SFLBSA	3'Sialyl-3-fucosyllactose-BSA
H2BSA	H Type II-APE-BSA
BGABSA	Blood Group A-BSA
BGBBSA	Blood Group B-BSA
GGGNHSA	Gala1,3Galb1,4GlcNAc-HSA
Ga3GBSA	Gala1,3Gal-BSA
Gb4GBSA	Galb1,4GalBSA
Ga2GBSA	Gala1,2GalBSA
4APBSA	4AP-BSA
LNFPIBSA	Lacto-N-fucopentaose I-BSA
LNFPIIBSA	Lacto-N-fucopentaose II-BSA
LNFPIIIBSA	Lacto-N-fucopentaose III-BSA
LNDHIBSA	Lacto-N-difucohexaose I-BSA
LebBSA	LNDI-BSA/ Lewis b-BSA
LexBSA	Lewis x-BSA
DiLexBSA	Di-Lex-APE-BSA
DiLexHSA	Di-Lewisx-APE-HSA
3LexHSA	Tri-Lex-APE-HSA
3SLexBSA3	3'Sialyl Lewis x-BSA

SLexBSA14	3'Sialyl Lewis x-BSA
6SuLexBSA	6-Sulfo Lewis x-BSA
6SuLeaBSA	6-Sulfo Lewis a-BSA
3SuLeaBSA	3-Sulfo Lewis a-BSA
3SuLexBSA	3-Sulfo Lewis x-BSA
DFPLNHHSA	Difucosyl-para-lacto-N-hexaose-APD-HSA
LeyHSA	Lewis y-tetrasaccharide-APE-HSA
3FLeyHSA	Tri-fucosyl-Ley-heptasaccharide-APE-HSA
LNnTHSA	Lacto-N-neotetraose-APD-HSA
LNTHSA	Lacto-N-tetraose-APD-HSA
SLNFVHSA	Sialyl-LNF V-APD-HSA
MMLNnHHSA	Monofucosyl, monosialyllacto-N-neohexaose-APD-HSA
SLNnTHSA	Sialyl-LNnT-penta-APD-HSA
GM1HSA	GM1-pentasaccharide-APD-HSA
aGM1HSA	Asialo-GM1-tetrasaccharide-APD-HSA
GlobNTHSA	Globo-N-tetraose-APD-HSA
GlobTHSA	Globotriose-APD-HSA
Inv	Invertase
Fibrin	Fibrinogen
A1AT	alpha-1-antitrypsin
Cerulo	Ceruloplasmin
AGP	alpha-1-acid glycoprotein
LacNAcaBSA	LacNAc-a-4AP-BSA
LacNAcb4APBSA	LacNAc-b-4AP-BSA
H2HSA	H-Type 2-APE-HSA
Ovomuc	Ovomucoid
RhaBSA	L-Rhamnose-Sp14-BSA
XManaBSA	Man-a-ITC-BSA
XLacbBSA	Lac-b-4AP-BSA
XGalbBSA	Gal-b-ITC-BSA
XylbBSA	Xyl-b-4AP-BSA
XylaBSA	Xyl-a-4AP-BSA
XGlcbBSA	Glc-b-4AP-BSA
FucaBSA	Fuc-a-4AP-BSA
FucbBSA	Fuc-b-4AP-BSA
GlcbITCBSA	Glc-b-ITC-BSA

Galb4APBSA	Gal-b-4AP-BSA
Neu5GcBSA	Neu5Gc-a-4AP-BSA
CollagenIV	Collagen type IV
D-GlobTHSA	Globotriose-HSA